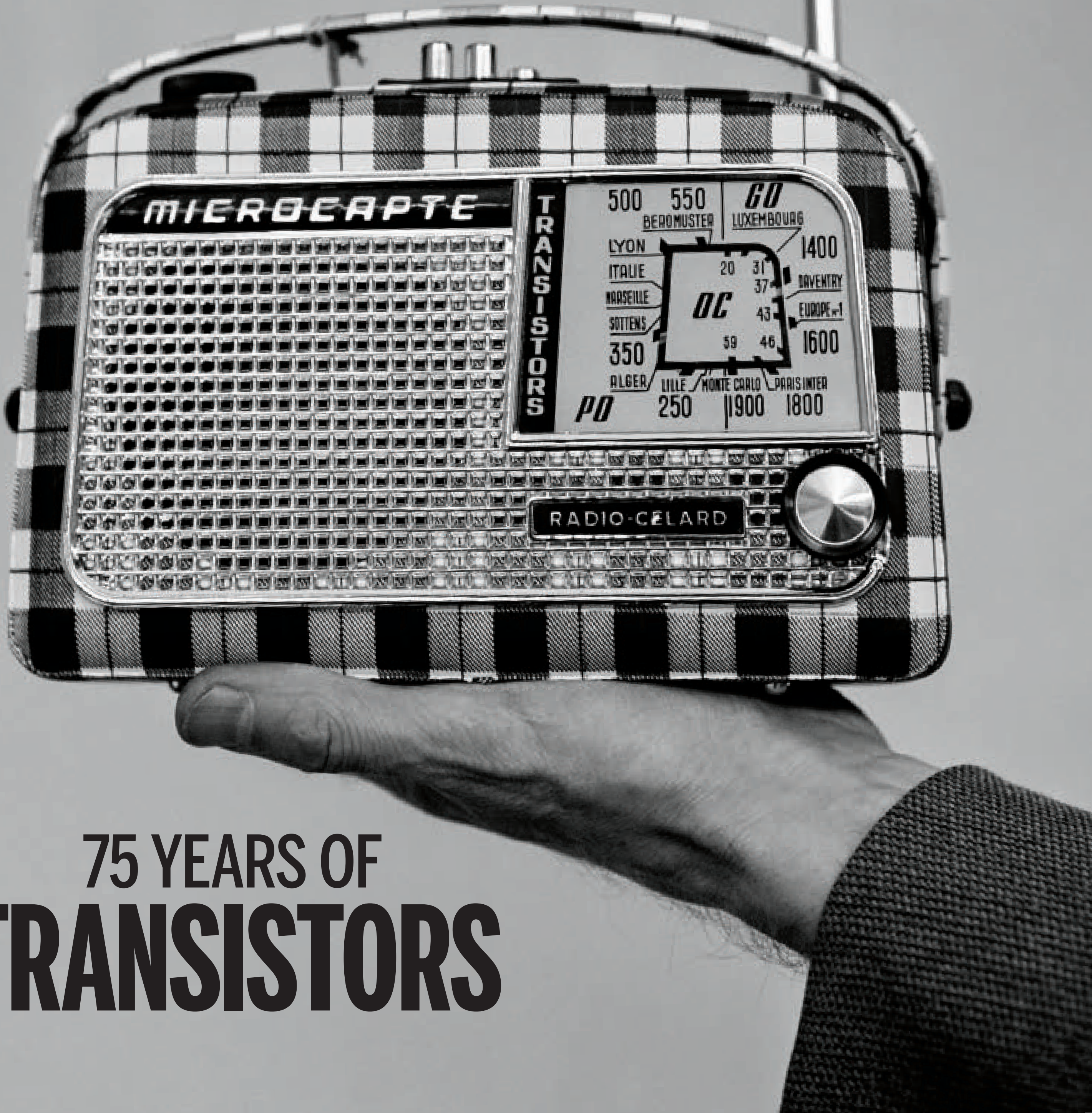
A Texas-size defense against storms and sea-level rise p. 698

Short-lived menstruation changes after COVID-19 vaccination p. 704

Phloem flow controls root outgrowth p. 762

Science

$15
18 NOVEMBER 2022
SPECIAL ISSUE
science.org
AAAS



75 YEARS OF TRANSISTORS

VISIT…

LANZAROTE Caliente.com

science.org/journal/sciimmunol



READY TO PUT THE SPOTLIGHT ON YOUR RESEARCH?

Submit your research:
cts.ScienceMag.org

Science Immunology
AAAS

Twitter: @SciImmunology
Facebook: @ScienceImmunology





Zhejiang Lab's research team members discuss chip design and testing plans.

Intelligent computing connects everything

At Zhejiang Lab, scientists develop new tools and techniques that will expand our understanding of the universe, improve health care, and open up new avenues of research.

The Internet of Everything (IoE) will connect people, processes, and devices. Scientists at Zhejiang Lab in Hangzhou, China see the IoE as a symbol of the new era of digital civilization, with intelligent computing encompassing computing methods and systems complemented by other technical capabilities that will drive the IoE. At Zhejiang Lab, scientists work on chips, algorithms, software, and novel computing architectures for intelligent computing.

""We take advantage of intelligent computing and novel research paradigms to revolutionarily advance the research in pertinent disciplines and develop new research models," says Tuo Shi, research specialist at Zhejiang Lab's Research Center for Intelligent Computing Hardware.

To drive this research, Zhejiang Lab created its Intelligent Computing Data Reactor. This facility collects virtually any type of data in real time and analyzes it with intelligent computing. As Shi explains, the reactor "efficiently uses intelligent-computing clusters, intelligent supercomputers, brain-like computers, as well as intelligent-computing software and hardware."

That collection of intelligent-computing technology can be applied to a wide range of fields including astronomy, genetics, materials and medical treatments.

Fostering FAST advances

In Guizhou, a mountainous region in southwest China, scientists built a radio telescope known as FAST (Five-hundred-meter Aperture Spherical radio Telescope). With a diameter of 500 meters, which is more than 1.5 times the height of the Eiffel Tower, FAST is the world's largest single-dish radio telescope, but that's just the beginning. As Shi notes, "The future 'FAST array' will be a large-scale telescope cluster consisting of up to five FASTs." This array will be used to explore the early universe, detect gravitational waves, and more.

To analyze data from projects like FAST, Zhejiang Lab developed its Astronomical Big Data Intelligent Computing and Service Platform. "The service platform takes the computing facility and intelligent technology as the base, and then uses data, algorithms, models, and knowledge to build a public knowledge base, and six engines, which include operation management, collaborative computing, knowledge construction, simulation deduction, data processing, and human–computer interaction," Shi says. "Driven by the intelligent data reactor engine, it schedules different computing tasks with the optimal computing resources, adapts the best computing methods, and forms the optimal results."

For FAST and the future FAST array, the Astronomical Big Data Intelligent Computing and Service Platform will support, as several examples, discoveries of pulsars, fast radio bursts, and extragalactic galaxies. It's already been used to propose a unified mechanism that can explain the evolution of the polarization frequency of repeated fast radio bursts.

In addition, this facility at Zhejiang Lab will advance other fields. As Shi says, "It will provide new methods, new tools, and new paradigms for scientific discovery, social governance, and the digital economy, and promote work in health care and other fields."

Making a material genome

"Although many people know about the genetic genome and other 'omes, the material genome refers to the acceleration of the discovery and design of novel materials using high value datasets and computational technologies," Shi explains. "Based on reliable experimental results obtained from characterization platforms and their databases involving chemical compositions, crystal structures, and physical properties, the relation between a material's structure and device performance can be established via theoretical predictions and simulations, leading to the discovery of new materials for innovative commercial purposes."

Zhejiang Lab's Material Intelligent Computing Platform combines big data, artificial intelligence, and domain knowledge to develop new approaches to creating intelligent materials. For example, artificial proteins can be developed for advanced cancer treatments and other disease treatments.

The future of connecting people and technology will change how we see and interact with the world. Intelligent computing will create the foundation of those advances.

Sponsored by

之江实验室
ZHEJIANG LAB

PHOTOS: **LEFT:** © ZHENGZAISHURU/SHUTTERSTOCK.COM; **RIGHT:** PROVIDED BY ZHEJIANG LAB

CONTENTS

18 NOVEMBER 2022 • VOLUME 378 • ISSUE 6621



Houston's petrochemical plants are at risk from hurricane-driven flooding.

NEWS

IN BRIEF

690 News at a glance

IN DEPTH

692 Science community braces for divided government
Split control of Congress could produce fiery hearings and limit new funding *By J. Mervis*

693 Tumors can teem with microbes. But what are they doing there?
New study suggests microbiomes can promote cancer by suppressing immune response and seeding metastases *By G. Sinha*

694 Whistleblower finds possible misconduct in his own papers
Matthew Schrag confronts a mentor after their joint work is flagged on the PubPeer website *By C. Piller*

696 Booming trade in mammoth ivory may be bad news for elephants
Paleontologists are urged to take a stand against a market that may provide cover for continued poaching *By M. Price*
PODCAST

697 Billionaire's crypto company collapse strands scientists
As FTX files for bankruptcy, the grantees its foundations supported may not see all of their pledged money *By R. F. Service*

FEATURES

698 Shelter from the storm
A plan to wall off Houston and nearby industry from flooding caused by hurricanes will cost tens of billions of dollars. Will it be enough?
By W. Cornwall

INSIGHTS

PERSPECTIVES

704 COVID-19 vaccination and menstruation
COVID-19 vaccination causes small changes to menstruation that quickly resolve
By V. Male

706 Defining the onset of the Anthropocene
Twelve sites are considered for defining the Anthropocene geological epoch
By C. N. Waters and S. D. Turner

709 Potential bias in genetic correlations
Mating patterns across two traits can inflate estimates of genetic overlap
By A. D. Grotzinger and M. C. Keller
RESEARCH ARTICLE p. 754

710 Adding functions to pyridines
Chemical reactions break a pyridine ring to allow its modification *By J. M. Joo*
REPORTS pp. 773 & 779

712 Using mirrors to control molecular dynamics
An optical cavity mixes molecular vibrations with light and changes chemical reactivity
By L. Chuntonov
REPORT p. 790

POLICY FORUM

713 Research under China's personal information law
The new law may present obstacles to some kinds of research *By X. Li* et al.

BOOKS *ET AL.*

716 The Gaia letters
Contextualized correspondence traces the emergence of a provocative hypothesis
By P. Falkowski

PHOTO: WITOLD SKRYPCZAK/ALAMY STOCK PHOTO

717 Aquatic life and its parallels
Blurring the personal and the scientific, an author probes the beauty, terror, and richness of the natural world *By J. Berwald*

LETTERS

718 Editorial Retraction
By H. H. Thorp

718 Brazil's preventable bridge disasters
By W. E. Magnusson

718 Welcoming Taiwan's diaspora scientists
By J. Huang

719 Life in Science: The elephant in the inbox
By C. T. A. Lewis

RESEARCH

IN BRIEF

741 From *Science* and other journals

RESEARCH ARTICLES

744 Developmental biology
Hippo signaling instructs ectopic but not normal organ growth *W. Kowalczyk* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABG3679

745 Human fertility
The mechanism of acentrosomal spindle assembly in human oocytes *T. Wu* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABQ7361

746 Cancer
Aberrant hyperexpression of the RNA binding protein FMRP in tumors mediates immune evasion *Q. Zeng* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABL7207

SPECIAL SECTION
75 YEARS OF TRANSISTORS

INTRODUCTION

720 From one transistor...

PERSPECTIVES

722 Moore's law: The journey ahead
M. S. Lundstrom and M. A. Alam

724 Toward gallium oxide power electronics *M. J. Tadjer*

REVIEWS

726 Carbon nanotube transistors: Making electronics from molecules
A. D. Franklin et al.

733 Toward attoJoule switching energy in logic transistors *S. Datta* et al.

SEE ALSO EDITORIAL p. 683

ON THE COVER

A hand holds a portable radio originally presented at the 21st Radio and Television Show at the Porte de Versailles Exhibition Center in Paris, France, on 9 September 1959. The invention of transistors 75 years ago spurred innovations that affected communication, computation, control systems, instrumentation, and other technologies. Today, the development of new transistor technologies continues on many fronts. See the special section beginning on page 720. *Photo: Keystone-France/Gamma-Rapho via Getty Images*



747 Solar cells
Initializing film homogeneity to retard phase segregation for stable perovskite solar cells
Y. Bai et al.

754 Human genomics
Cross-trait assortative mating is widespread and inflates genetic correlation estimates
R. Border et al.
PERSPECTIVE p. 709

762 Plant science
Hydraulic flux–responsive hormone redistribution determines root branching
P. Mehra et al.

REPORTS

768 3D printing
Mechanical nanolattices printed using nanocluster-based photoresists *Q. Li* et al.

Organic chemistry
773 Halogenation of the 3-position of pyridines through Zincke imine intermediates *B. T. Boyle* et al.
779 Radical and ionic *meta*-C–H functionalization of pyridines, quinolines, and isoquinolines *H. Cao* et al.
PERSPECTIVE p. 710

785 Quantum physics
Noise-resilient edge modes on a chain of superconducting qubits
X. Mi et al.

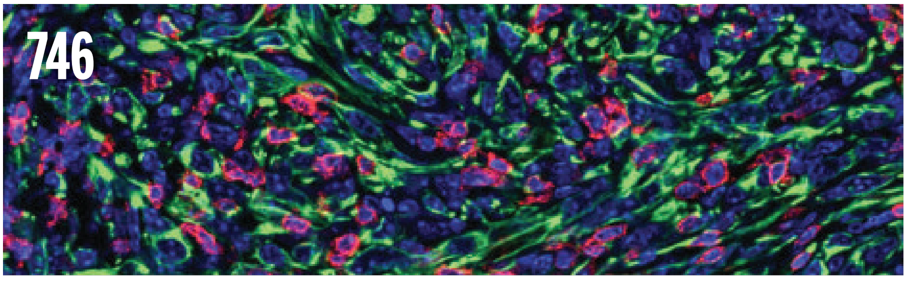
790 Chemical physics
Cavity-enabled enhancement of ultrafast intramolecular vibrational redistribution over pseudorotation *T.-T. Chen* et al.
PERSPECTIVE p. 712

DEPARTMENTS

683 Editorial
Shockley was a racist and eugenicist
By H. H. Thorp
75 YEARS OF TRANSISTORS SECTION p. 720

798 Working Life
Silent no longer *By W. P. Suza*

Science Staff 682
Science Careers 796



Killer T cells (red) attack susceptible cancer cells (green).

IMAGE: JEREMY GUILLOT

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2022 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2212; Foreign postage extra: Air assist delivery: $98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.science.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

Science AAAS

Editor-in-Chief Holden Thorp, hthorp@aaas.org

Executive Editor Valda Vinson

Editor, Research Jake S. Yeston **Editor, Insights** Lisa D. Chong **Managing Editor** Lauren Kmec

DEPUTY EDITORS Gemma Alderton (UK),Stella M. Hurtley (UK), Phillip D. Szuromi, Sacha Vignieri **SR. EDITORS** Caroline Ash (UK), Michael A. Funk, Brent Grocholski, Pamela J. Hines, Di Jiang, Priscilla N. Kelly, Marc S. Lavine (Canada), Mattia Maroso, Yevgeniya Nusinovich, Ian S. Osborne (UK), L. Bryan Ray, Seth Thomas Scanlon (UK), H. Jesse Smith, Keith T. Smith (UK), Jelena Stajic, Peter Stern (UK), Valerie B. Thompson, Brad Wible **ASSOCIATE EDITORS** Bianca Lopez, Corinne Simonti, Yury V. Suleymanov, Ekeoma Uzogara **LETTERS EDITOR** Jennifer Sills **LEAD CONTENT PRODUCTION EDITORS** Chris Filiatreau, Harry Jach **SR. CONTENT PRODUCTION EDITOR** Amelia Beyna **CONTENT PRODUCTION EDITORS** Julia Haber-Katris, Nida Masiulis, Abigail Shashikanth, Suzanne M. White **SR. EDITORIAL MANAGERS** Carolyn Kyle, Beverly Shields **SR. PROGRAM ASSOCIATE** Maryrose Madrid **EDITORIAL ASSOCIATE** Joi S. Granger **SR. EDITORIAL COORDINATORS** Aneera Dobbins, Jeffrey Hearn, Lisa Johnson, Jerry Richardson, Hilary Stewart (UK), Alice Whaley (UK), Anita Wynn **EDITORIAL COORDINATORS** Maura Byrne, Alexander Kief, Ronmel Navas, Isabel Schnaidt, Qiyam Stewart, Brian White **RESEARCH & DATA ANALYST** Jessica L. Slater **ASI DIRECTOR, OPERATIONS** Janet Clements (UK) **ASI SR. OFFICE ADMINISTRATOR** Jessica Waldock (UK)

News Editor Tim Appenzeller

NEWS MANAGING EDITOR John Travis **INTERNATIONAL EDITOR** Martin Enserink **DEPUTY NEWS EDITORS** Shraddha Chakradhar, Elizabeth Culotta, Lila Guterman, David Grimm, Eric Hand (Europe), David Malakoff **SR. CORRESPONDENTS** Daniel Clery (UK), Jon Cohen, Jeffrey Mervis, Elizabeth Pennisi **ASSOCIATE EDITORS** Jeffrey Brainard, Michael Price, Kelly Servick **NEWS REPORTERS** Adrian Cho, Jennifer Couzin-Frankel, Jocelyn Kaiser, Rodrigo Pérez Ortega (Mexico City), Robert F. Service, Erik Stokstad, Paul Voosen, Meredith Wadman **INTERNS** Zack Savitsky, Viviana Flores, Katherine Irving **CONTRIBUTING CORRESPONDENTS** Warren Cornwall, Andrew Curry (Berlin), Ann Gibbons, Sam Kean, Eli Kintisch, Kai Kupferschmidt (Berlin), Andrew Lawler, Mitch Leslie, Eliot Marshall, Virginia Morell, Dennis Normile (Tokyo), Elisabeth Pain (Careers), Charles Piller, Gabriel Popkin, Michael Price, Joshua Sokol, Richard Stone, Emily Underwood, Gretchen Vogel (Berlin), Lizzie Wade (Mexico City) **CAREERS** Rachel Bernstein (Editor), Katie Langin (Associate Editor) **COPY EDITORS** Julia Cole (Senior Copy Editor), Morgan Everett, Cyra Master (Copy Chief) **ADMINISTRATIVE SUPPORT** Meagan Weiland

Creative Director Beth Rakouskas

DESIGN MANAGING EDITOR Chrystal Smith **GRAPHICS MANAGING EDITOR** Chris Bickel **PHOTOGRAPHY MANAGING EDITOR** Emily Petersen **MULTIMEDIA MANAGING PRODUCER** Kevin McLean **WEB CONTENT STRATEGY MANAGER** Kara Estelle-Powers **DESIGN EDITOR** Marcy Atarod **DESIGNER** Christina Aycock **SENIOR SCIENTIFIC ILLUSTRATOR** Valerie Altounian **SCIENTIFIC ILLUSTRATORS** Austin Fisher, Kellie Holoski, Ashley Mastin **INTERACTIVE GRAPHICS EDITOR** Kelly Franklin **SENIOR GRAPHICS SPECIALISTS** Holly Bishop, Nathalie Cary **SENIOR PHOTO EDITOR** Charles Borst **PHOTO EDITOR** Kaitlyn Dolan **SENIOR PODCAST PRODUCER** Sarah Crespi **VIDEO PRODUCER** Meagan Cantwell **SOCIAL MEDIA STRATEGIST** Jessica Hubbard **SOCIAL MEDIA PRODUCER** Sabrina Jenkins **WEB DESIGNER** Jennie Pajerowski

Chief Executive Officer and Executive Publisher Sudip Parikh

Publisher, Science Family of Journals Bill Moran

DIRECTOR, BUSINESS SYSTEMS AND FINANCIAL ANALYSIS Randy Yi **DIRECTOR, BUSINESS OPERATIONS & ANALYSIS** Eric Knott **DIRECTOR OF ANALYTICS** Enrique Gonzales **MANAGER, BUSINESS OPERATIONS** Jessica Tierney **MANAGER, BUSINESS ANALYSIS** Cory Lipman **BUSINESS ANALYSTS** Kurt Ennis, Maggie Clark **FINANCIAL ANALYST** Isacco Fusi **BUSINESS OPERATIONS ADMINISTRATOR** Taylor Fisher **SENIOR PRODUCTION MANAGER** Jason Hillman **SENIOR MANAGER, PUBLISHING AND CONTENT SYSTEMS** Marcus Spiegler **CONTENT OPERATIONS MANAGER** Rebecca Doshi **SENIOR CONTENT & PUBLISHING SYSTEMS SPECIALIST** Jacob Hedrick **SENIOR PRODUCTION SPECIALIST** Kristin Wowk **PRODUCTION SPECIALISTS** Kelsey Cartelli, Audrey Diggs **DIGITAL PRODUCTION MANAGER** Lisa Stanford **CONTENT SPECIALIST** Kimberley Oster **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **DESIGNER, CUSTOM PUBLISHING** Jeremy Huntsinger **SR. TRAFFIC ASSOCIATE** Christine Hall **SPECIAL PROJECTS ASSOCIATE** Sarah Dhere

ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT Justin Sawyers **GLOBAL MARKETING MANAGER** Allison Pritchard **DIGITAL MARKETING MANAGER** Aimee Aponte **JOURNALS MARKETING MANAGER** Shawana Arnold **MARKETING ASSOCIATES** Aaron Helmbrecht, Ashley Hylton, Mike Romano, Lorena Chirinos Rodriguez, Jenna Voris **SENIOR DESIGNER** Kim Huynh

DIRECTOR AND SENIOR EDITOR, CUSTOM PUBLISHING Sean Sanders **ASSISTANT EDITOR, CUSTOM PUBLISHING** Jackie Oberst **PROJECT MANAGER** Melissa Collins

DIRECTOR, PRODUCT & PUBLISHING DEVELOPMENT Chris Reid **DIRECTOR, BUSINESS STRATEGY AND PORTFOLIO MANAGEMENT** Sarah Whalen **ASSOCIATE DIRECTOR, PRODUCT MANAGMENT** Kris Bishop **PRODUCT DEVELOPMENT MANAGER** Scott Chernoff **PUBLISHING TECHNOLOGY MANAGER** Michael Di Natale **SR. PRODUCT ASSOCIATE** Robert Koepke **PRODUCT ASSOCIATE** Caroline Breul, Anne Mason **SPJ ASSOCIATE MANAGER** Samantha Bruno Fuller **SPJ ASSOCIATE** Casey Buchta

MARKETING MANAGER Kess Knight **BUSINESS DEVELOPMENT MANAGER** Rasmus Andersen **SENIOR INSTITUTIONAL LICENSING MANAGER** Ryan Rexroth **INSTITUTIONAL LICENSING MANAGER** Marco Castellan, Claudia Paulsen-Young **SENIOR MANAGER, INSTITUTIONAL LICENSING OPERATIONS** Judy Lillibridge **SENIOR OPERATIONS ANALYST** Lana Guz **SYSTEMS & OPERATIONS ANALYST** Ben Teincuff **FULFILLMENT ANALYST** Aminta Reyes

DIRECTOR, GLOBAL SALES Tracy Holmes **US EAST COAST AND MID WEST SALES** Stephanie O'Connor **US MID WEST, MID ATLANTIC AND SOUTH EAST SALES** Chris Hoag **US WEST COAST SALES** Lynne Stickrod **ASSOCIATE DIRECTOR, ROW** Roger Goncalves **SALES REP, ROW** Sarah Lelarge **SALES ADMIN ASSISTANT, ROW** Victoria Glasbey **DIRECTOR OF GLOBAL COLLABORATION AND ACADEMIC PUBLISHING RELATIONS, ASIA** Xiaoying Chu **ASSOCIATE DIRECTOR, INTERNATIONAL COLLABORATION** Grace Yao **SALES MANAGER** Danny Zhao **MARKETING MANAGER** Kilo Lan **ASCA CORPORATION, JAPAN** Rie Rambelli (Tokyo), Miyuki Tani (Osaka)

DIRECTOR, COPYRIGHT, LICENSING AND SPECIAL PROJECTS Emilie David **RIGHTS AND PERMISSIONS ASSOCIATE** Elizabeth Sandler **LICENSING ASSOCIATE** Virginia Warren **CONTRACT SUPPORT SPECIALIST** Michael Wheeler

MAIN HEADQUARTERS
Science/AAAS
1200 New York Ave. NW
Washington, DC 20005

SCIENCE INTERNATIONAL
Clarendon House
Clarendon Road
Cambridge, CB2 8FH, UK

SCIENCE CHINA
Room 1004, Culture Square
No. 59 Zhongguancun St.
Haidian District, Beijing, 100872

SCIENCE JAPAN
ASCA Corporation
Sibaura TY Bldg. 4F, 1-14-5
Shibaura Minato-ku
Tokyo, 108-0073 Japan

EDITORIAL
science_editors@aaas.org

NEWS
science_news@aaas.org

INFORMATION FOR AUTHORS
science.org/authors/
science-information-authors

REPRINTS AND PERMISSIONS
science.org/help/
reprints-and-permissions

MEDIA CONTACTS
scipak@aaas.org

MULTIMEDIA CONTACTS
SciencePodcast@aaas.org
ScienceVideo@aaas.org

INSTITUTIONAL SALES AND SITE LICENSES
science.org/librarian

PRODUCT ADVERTISING & CUSTOM PUBLISHING
advertising.science.org/
products-services
science_advertising@aaas.org

CLASSIFIED ADVERTISING
advertising.science.org/
science-careers
advertise@sciencecareers.org

JOB POSTING CUSTOMER SERVICE
employers.sciencecareers.org
support@sciencecareers.org

MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS
science.org/subscriptions

MEMBER BENEFITS
aaas.org/membership/benefits

AAAS BOARD OF DIRECTORS
CHAIR Susan G. Amara
PRESIDENT Gilda A. Barabino
PRESIDENT-ELECT Keith Yamamoto
TREASURER Carolyn N. Ainslie
CHIEF EXECUTIVE OFFICER
Sudip Parikh
BOARD Cynthia M. Beall
Ann Bostrom
Janine Austin Clayton
Kaye Husbands Fealing
Maria M. Klawe
Jane Maienschein
Robert B. Millard
Babak Parviz
William D. Provine
Juan S. Ramírez Lugo
Susan M. Rosenberg

Science serves as a forum for discussion of important issues related to the advancement of science by publishing material on which a consensus has been reached as well as including the presentation of minority or conflicting points of view. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

BOARD OF REVIEWING EDITORS (Statistics board members indicated with **S**)

Erin Adams, *U. of Chicago*
Takuzo Aida, *U. of Tokyo*
Leslie Aiello, *Wenner-Gren Fdn.*
Deji Akinwande, *UT Austin*
Judith Allen, *U. of Manchester*
Marcella Alsan, *Harvard U.*
James Analytis, *UC Berkeley*
Paola Arlotta, *Harvard U.*
Delia Baldassarri, *NYU*
Nenad Ban, *ETH Zürich*
Christopher Barratt, *U. of Dundee*
Nandita Basu, *U. of Waterloo*
Franz Bauer, *Pontificia U. Católica de Chile*
Ray H. Baughman, *UT Dallas*
Carlo Beenakker, *Leiden U.*
Yasmine Belkaid, *NIAID, NIH*
Philip Benfey, *Duke U.*
Kiros T. Berhane, *Columbia U.*
Joseph J. Berry, *NREL*
Alessandra Biffi, *Harvard Med.*
Chris Bowler, *École Normale Supérieure*
Ian Boyd, *U. of St. Andrews*
Malcolm Brenner, *Baylor Coll. of Med.*
Emily Brodsky, *UC Santa Cruz*
Ron Brookmeyer, *UCLA* (**S**)
Christian Büchel, *UKE Hamburg*
Dennis Burton, *Scripps Res.*
Carter Tribley Butts, *UC Irvine*
György Buzsáki, *NYU School of Med.*
Mariana Byndloss, *Vanderbilt U. Med. Ctr.*
Annmarie Carlton, *UC Irvine*
Simon Cauchemez, *Inst. Pasteur*
Ling-Ling Chen, *SIBCB, CAS*
Wendy Cho, *UIUC*
Ib Chorkendorff, *Denmark TU*
Karlene Cimprich, *Stanford U.*
James J. Collins, *MIT*
Robert Cook-Deegan, *Arizona State U.*
Virginia Cornish *Columbia U.*
Carolyn Coyne, *Duke U.*
Roberta Croce, *VU Amsterdam*
Christina Curtis, *Stanford U.*
Ismaila Dabo, *Penn State U.*
Jeff L. Dangl, *UNC*
Chiara Daraio, *Caltech*
Nicolas Dauphas, *U. of Chicago*
Frans de Waal, *Emory U.*
Claude Desplan, *NYU*
Sandra Díaz, *U. Nacional de Córdoba*
Samuel Díaz-Muñoz, *UC Davis*
Ulrike Diebold, *TU Wien*
Stefanie Dimmeler, *Goethe-U. Frankfurt*
Hong Ding, *Inst. of Physics, CAS*
Dennis Discher, *UPenn*
Jennifer A. Doudna, *UC Berkeley*
Ruth Drdla-Schutting, *Med. U. Vienna*
Raissa M. D'Souza, *UC Davis*
Bruce Dunn, *UCLA*
William Dunphy, *Caltech*
Scott Edwards, *Harvard U.*
Todd Ehlers, *U. of Tübingen*
Nader Engheta, *UPenn*
Karen Ersche, *U. of Cambridge*
Beate Escher, *UFZ & U. of Tübingen*
Barry Everitt, *U. of Cambridge*
Vanessa Ezenwa, *U. of Georgia*
Toren Finkel, *U. of Pitt. Med. Ctr.*
Gwenn Flowers, *Simon Fraser U.*
Natascha Förster Schreiber, *MPI Extraterrestrial Phys.*
Peter Fratzl, *MPI Potsdam*
Elaine Fuchs, *Rockefeller U.*
Jay Gallagher, *U. of Wisconsin*
Daniel Geschwind, *UCLA*
Ramon Gonzalez, *U. of South Florida*
Sandra González-Bailón, *UPenn*
Gillian Griffiths, *U. of Cambridge*
Nicolas Gruber, *ETH Zürich*
Hua Guo, *U. of New Mexico*
Taekjip Ha, *Johns Hopkins U.*
Daniel Haber, *Mass. General Hos.*
Sharon Hammes-Schiffer, *Yale U.*
Wolf-Dietrich Hardt, *ETH Zürich*
Louise Harra, *UCL*
Carl-Philipp Heisenberg, *IST Austria*
Janet G. Hering, *Eawag*
Christoph Hess, *U. of Basel & U. of Cambridge*
Heather Hickman, *NIAID, NIH*
Hans Hilgenkamp, *U. of Twente*
Janneke Hille Ris Lambers, *ETH Zürich*
Kai-Uwe Hinrichs, *U. of Bremen*
Deirdre Hollingsworth, *U. of Oxford*
Randall Hulet, *Rice U.*
Auke Ijspeert, *EPFL*
Gwyneth Ingram, *ENS Lyon*
Darrell Irvine, *MIT*
Akiko Iwasaki, *Yale U.*
Stephen Jackson, *USGS & U. of Arizona*
Erich Jarvis, *Rockefeller U.*
Peter Jonas, *IST Austria*
Johanna Joyce, *U. de Lausanne*
Matt Kaeberlein, *U. of Wash.*
William Kaelin Jr., *Dana-Farber*
Daniel Kammen, *UC Berkeley*
Kisuk Kang, *Seoul Nat. U.*
Sabine Kastner, *Princeton U.*
V. Narry Kim, *Seoul Nat. U.*
Robert Kingston, *Harvard Med.*
Nancy Knowlton, *Smithsonian Institution*
Etienne Koechlin, *École Normale Supérieure*
Alex L. Kolodkin, *Johns Hopkins U.*
LaShanda Korley, *U. of Delaware*
Julija Krupic, *U. of Cambridge*
Paul Kubes, *U. of Calgary*
Chris Kuzawa, *Northwestern U.*
Laura Lackner, *Northwestern U.*
Gabriel Lander, *Scripps Res.* (**S**)
Mitchell A. Lazar, *UPenn*
Hedwig Lee, *Duke U.*
Luis Liz-Marzán, *CIC biomaGUNE*
Omar Lizardo, *UCLA*
Jonathan Losos, *Wash. U. in St. Louis*
Ke Lu, *Inst. of Metal Res., CAS*
Christian Lüscher, *U. of Geneva*
Jean Lynch-Stieglitz, *Georgia Inst. of Tech.*
David Lyons, *U. of Edinburgh*
Fabienne Mackay, *QIMR Berghofer*
Zeynep Madak-Erdogan, *UIUC*
Anne Magurran, *U. of St. Andrews*
Ari Pekka Mähönen, *U. of Helsinki*
Asifa Majid, *U. of Oxford*
Oscar Marín, *King's Coll. London*
Charles Marshall, *UC Berkeley*
Christopher Marx, *U. of Idaho*
David Masopust, *U. of Minnesota*
Geraldine Masson, *CNRS*
C. Robertson McClung, *Dartmouth*
Rodrigo Medellín, *U. Nacional Autónoma de México*
C. Jessica Metcalf, *Princeton U.*
Baoxia Mi, *UC Berkeley*
Tom Misteli, *NCI, NIH*
Alison Motsinger-Reif, *NIEHS, NIH* (**S**)
Danielle Navarro, *U. of New South Wales*
Daniel Nettle, *Newcastle U.*
Daniel Neumark, *UC Berkeley*
Beatriz Noheda, *U. of Groningen*
Helga Nowotny, *Vienna Sci. & Tech. Fund*
Pilar Ossorio, *U. of Wisconsin*
Andrew Oswald, *U. of Warwick*
Isabella Pagano, *Istituto Nazionale di Astrofisica*
Elizabeth Levy Paluck, *Princeton U.*
Jane Parker, *MPI Cologne*
Giovanni Parmigiani, *Dana-Farber* (**S**)
Daniel Pauly, *U. of British Columbia*
Ana Pêgo, *U. do Porto*
Samuel Pfaff, *Salk Inst.*
Julie Pfeiffer, *UT Southwestern Med. Ctr.*
Philip Phillips, *UIUC*
Matthieu Piel, *Inst. Curie*
Kathrin Plath, *UCLA*
Martin Plenio, *Ulm U.*
Katherine Pollard, *UCSF*
Elvira Poloczanska, *Alfred-Wegener-Inst.*
Julia Pongratz, *Ludwig Maximilians U.*
Philippe Poulin, *CNRS*
Lei Stanley Qi, *Stanford U.*
Trevor Robbins, *U. of Cambridge*
Joeri Rogelj, *Imperial Coll. London*
Amy Rosenzweig, *Northwestern U.*
Mike Ryan, *UT Austin*
Miquel Salmeron, *Lawrence Berkeley Nat. Lab*
Nitin Samarth, *Penn State U.*
Erica Ollmann Saphire, *La Jolla Inst.*
Joachim Saur, *U. zu Köln*
Alexander Schier, *Harvard U.*
Wolfram Schlenker, *Columbia U.*
Susannah Scott, *UC Santa Barbara*
Anuj Shah, *U. of Chicago*
Vladimir Shalaev, *Purdue U.*
Jie Shan, *Cornell U.*
Beth Shapiro, *UC Santa Cruz*
Jay Shendure, *U. of Wash.*
Steve Sherwood, *U. of New South Wales*
Brian Shoichet, *UCSF*
Robert Siliciano, *JHU School of Med.*
Lucia Sivilotti, *UCL*
Emma Slack, *ETH Zürich & U. of Oxford*
Richard Smith, *UNC* (**S**)
John Speakman, *U. of Aberdeen*
Tara Spires-Jones, *U. of Edinburgh*
Allan C. Spradling, *Carnegie Institution for Sci.*
V. S. Subrahmanian, *Northwestern U.*
Ira Tabas, *Columbia U.*
Eriko Takano, *U. of Manchester*
A. Alec Talin, *Sandia Natl. Labs*
Patrick Tan, *Duke-NUS Med. School*
Sarah Teichmann, *Wellcome Sanger Inst.*
Rocio Titiunik, *Princeton U.*
Shubha Tole, *Tata Inst. of Fundamental Res.*
Maria-Elena Torres Padilla, *Helmholtz Zentrum München*
Kimani Toussaint, *Brown U.*
Barbara Treutlein, *ETH Zürich*
Jason Tylianakis, *U. of Canterbury*
Wim van der Putten, *Netherlands Inst. of Ecology*
Matthew Vander Heiden, *MIT*
Ivo Vankelecom, *KU Leuven*
Judith Varner, *UC San Diego*
Henrique Veiga-Fernandes, *Champalimaud Fdn.*
Reinhilde Veugelers, *KU Leuven*
Bert Vogelstein, *Johns Hopkins U.*
Julia Von Blume, *Yale School of Med.*
David Wallach, *Weizmann Inst.*
Jane-Ling Wang, *UC Davis* (**S**)
Jessica Ware, *Amer. Mus. of Natural Hist.*
David Waxman, *Fudan U.*
Chris Wikle, *U. of Missouri* (**S**)
Terrie Williams, *UC Santa Cruz*
Ian A. Wilson, *Scripps Res.* (**S**)
Hao Wu, *Harvard U.*
Li Wu, *Tsinghua U.*
Wei Xie, *Tsinghua U.*
Benjamin Youngblood, *St. Jude*
Yu Xie, *Princeton U.*
Jan Zaanen, *Leiden U.*
Kenneth Zaret, *UPenn School of Med.*
Lidong Zhao, *Beihang U.*
Bing Zhu, *Inst. of Biophysics, CAS*
Xiaowei Zhuang, *Harvard U.*
Maria Zuber, *MIT*

EDITORIAL

Shockley was a racist and eugenicist



H. Holden Thorp
Editor-in-Chief, *Science* journals.
hthorp@aaas.org; @hholdenthorp

This week's issue on the 75th anniversary of the transistor describes a triumph of both basic and applied science. What started out as studies on the fundamental physics of silicon led to the device that makes it possible to read this article online. The coinventor of the transistor, William Shockley, who along with John Bardeen and Walter Brattain won the 1956 Nobel Prize in Physics, is correctly recognized as a primary architect of the computer age. Gordon Moore (cofounder of Intel Corporation) famously said that Shockley put the silicon in "Silicon Valley." Appallingly, Shockley devoted the latter part of his life to promoting racist views, arguing that higher IQs among Blacks were correlated with higher extents of Caucasian ancestry, and advocating for voluntary sterilization of Black women. At the time, *Science* did not condemn Shockley for what he was: a charlatan who used his scientific credentials to advance racist ideology.

The failure of *Science* to condemn Shockley began in 1968, when it published a letter lamenting the fact that he was prohibited from speaking at the Polytechnic Institute of Brooklyn. The letter repeated the familiar trope that Shockley was simply asking questions about the role of race in intelligence. But Shockley had no scientific basis for doing so, he was not submitting peer-reviewed papers on the topic, and most importantly, he was using his ideas as the basis for promoting eugenics. Such a debate had no place in this journal.

Shockley was part of a cadre of physicists who advanced ideas outside of their area of expertise to promote a right-wing agenda. He was a close friend of Frederick Seitz—president of both the National Academy of Sciences and Rockefeller University—who, following a career in physics, became a purveyor of misinformation on tobacco, nuclear weapons, and climate change. Like Shockley, Seitz carried out his nonphysics work through op-eds and conservative think tanks, not through the accepted mechanism of peer review that he used in doing physics. Seitz was not, at least publicly, as overtly in favor of eugenics as was Shockley, but he was a strong advocate for genetic determinism, even claiming at the behest of the cigarette industry that tobacco itself was not harmful because genetics determined whether smokers would ultimately contract lung cancer.

Following Shockley's death in 1989, *Nature* correctly called out his racism in an obituary, but then published a letter from Seitz defending Shockley and claiming that the reason Shockley became a eugenicist was because of physical trauma he experienced in a near-fatal car accident. When *Science* wrote about this dustup, it referred to Shockley's ideas as merely "unpopular" and "extremely controversial." It then ran a letter from an even more notorious eugenicist, J. Philippe Rushton, who argued that by merely covering the disagreement at *Nature*, *Science* was delivering an "ad hominem attack." In addition to an ill-advised decision to publish Rushton's letter, *Science* posted a response saying, "no criticism of Shockley was intended." Yikes.

> **"The process of science is one of continual revision, but it's also one that must have a conscience."**

Looking back, it's clear that what was intended as an attempt to make room for dissent and discussion only served to abet Shockley and his cohorts in their effort to build support for eugenics. *Science* gave them a platform and inadequate scorn. The lesson is that we at *Science* need to make more effort to think about everything that we do, not only from the standpoint of communicating science to the public, but also as an organization that above all, supports all of humanity. The process of science is one of continual revision, but it's also one that must have a conscience.

It was only a few months ago, in a commentary on racism in science by Ebony Omotola McGee, that Shockley was described in our pages in the terms he deserved. But as recently as 2001, *Science* described him simply as a "transistor inventor and race theorist." That won't cut it anymore. As of today, a link to this editorial will appear along with any mention of Shockley in this journal.

Make no mistake. Shockley was a racist. Shockley was a eugenicist. That's all.

–H. Holden Thorp

PHOTO: CAMERON DAVIDSON

10.1126/science.adf8117



PHOTO: © DMITRIY RYBIN/SHUTTERSTOCK.COM

The multi(dimension)verse of computing

Communications technology is changing the way we live at an ever-faster pace. Massive amounts of data can be transferred around the world in a flash, and mathematical problems that have eluded the best minds can now be solved in mere months by supercomputing. The arrival of quantum computing, however, will make these technologies archaic. If its expected potential is realized, then quantum computing will be as superior to supercomputing as the works of Shakespeare compared to the output of a monkey at a typewriter.

Supercomputing remains the pinnacle of digital computing; it works by taking classical bits of information and, with enough energy, making an extraordinary number of calculations to find a solution to complex problems. However, whereas a classical bit only takes one of two states (either a 0 or a 1), a qubit, the information basis of quantum computing, can represent a superposition of the two states. This distinction allows quantum computers to analyze multidimensional spaces to recognize patterns that simply cannot be observed by classical computers. Not only does this new perspective reduce the time needed to calculate a solution, but it also expands by leaps and bounds the types of problems that can be solved.

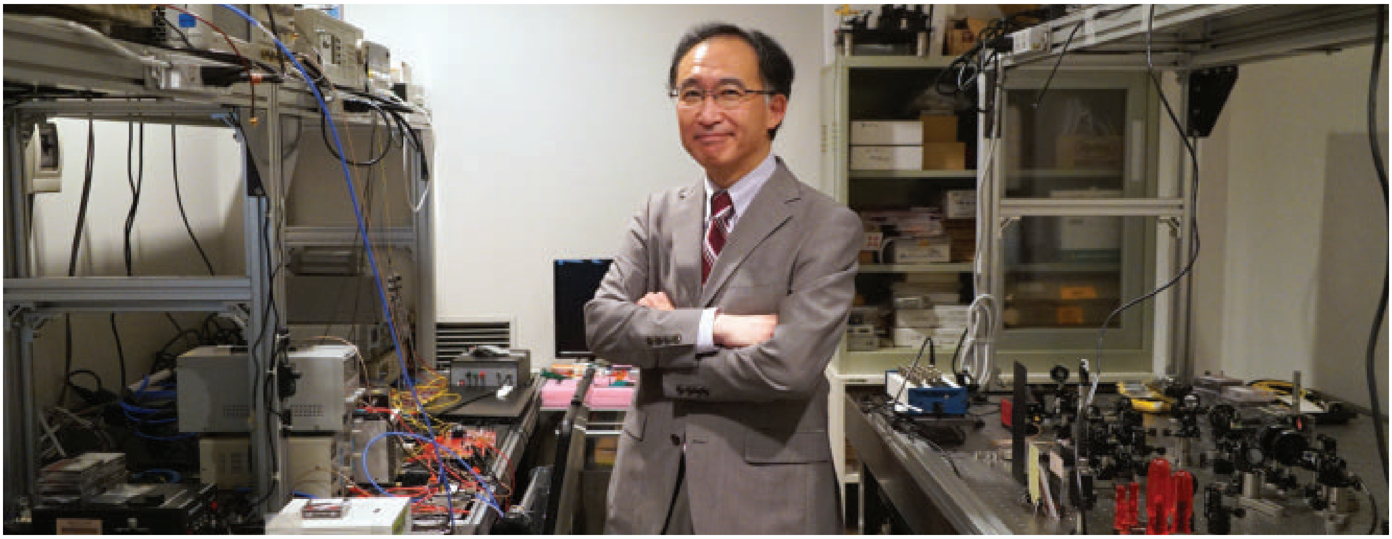
The applications are endless and will impact every area of our lives. These computers will be better at designing drugs, planning global supply chains, reducing financial risk, and predicting the weather. They could even reveal the elusive secrets of the universe.

Quantum computing may not get the same attention as the space race did many decades ago, but countries of all sizes recognize the implications of this technology and are competing to take the lead. Included in this list is Japan, which aims to announce four new quantum computing research centers by the end of the fiscal year. Japanese government funding agencies have invested more than USD $250 million in this field in the past 15 years, and Japanese industry is matching these amounts. Moreover, in 2018, the government announced its "Q-LEAP" initiative, committing hundreds of millions more into research on hardware and software related to quantum computing. All this is in anticipation of a market size that is estimated to increase almost 10-fold in the next 5 years.

The impressive contributions by Japanese government and industry are matched by the discoveries made through Japan's academia–industry collaborations. Building on its recognized strengths in semiconductors and high-performance computing, Japan has achieved a number of firsts in quantum computing. Quantum key distribution and quantum annealing are just two of the areas it excels in as it advances into the commercial market, pushed by companies such as Fujitsu, Toshiba, and NEC, among many others. Commercialization is also being facilitated by organizations like the Quantum ICT Forum, which is bringing the best minds in the field to translate technologies and guide policy.

Quantum computers will not be in commission for another decade or two because of the many functional challenges that remain, such as the short coherence time of qubits and the large error rates in quantum operations. Ironically, these problems will be solved iteratively, much in the way digital computing works. In the meantime, Japan is already producing quantum-inspired digital technologies that solve many of the combinatorial optimization problems eluding even the best supercomputers. Although not as fast as true quantum computers, these quantum-based devices can reduce problem-solving time from months to hours and from days to seconds. Moreover, they are available right now.

For decades, Japan has been a herald of the future. From its bullet trains to its robot-staffed hotels, the country has played a major role in advancing technological progress on the world stage. Now its next performance is about to debut, with a new star—quantum computing.



Director of the Quantum ICT Forum Akihisa Tomita is helping to usher in quantum technology.

PHOTO: PROVIDED BY QUANTUM ICT FORUM

Gathering the best minds in quantum technology

The strength of Japan's research in quantum information and communication technology (ICT) can be seen all across the country. In terms of R&D, Japan's universities are at the top of the field and its companies are at the forefront of developing key technologies such as superconducting qubits, quantum annealing, and optical lattice clocks. However, so long as this progress remains scattered, its benefits to society will be limited. It was in response to this problem that the Quantum ICT Forum was founded, to quicken Japan's adoption of quantum ICT.

History has shown that societies have gone through four great transformations, beginning with the hunter-gatherer society and moving through the agricultural, industrial, and most recently, information societies. Japan is now preparing for Society 5.0, which aims to connect cyberspace with physical space. This will only happen through a new generation of ICT, one that will depend heavily on quantum-based technologies.

"Our goal is to bring researchers and industry together so that we can use quantum technologies for a healthier, more productive society," says Akihisa Tomita, director of the Quantum ICT Forum.

However, unlike academic societies, the Forum does more than invite researchers to present their studies. It is designed to promote commercialization so that Japan stays at the forefront of ICT.

In particular, the Forum has identified three sectors—high-speed computers, highly sensitive and accurate sensors, and faster and more secure communications—as having the biggest impact on ICT in the near future. Thus, it has formed three committees: one on quantum computing, another on quantum sensing, and the last on quantum key distribution.

These committees serve many purposes that benefit academic, industrial, and public interests. First and foremost, they promote collaborative research. Second, they bring consensus to standards and provide guidance on government policy. Finally, they promote public awareness through information dissemination, whether in the form of exhibitions, seminars, newsletters, or other mediums.

"Each committee is looking at best practices to keep Japan at the leading edge of quantum ICT. The most important efforts involve translating research into valuable customer products and assisting with policy making so that society benefits," says Tomita.

Bringing more people to quantum ICT

Officially, the Quantum ICT Forum was founded in 2019, but its history can be traced to 2001, when Japanese researchers were rapidly making several major discoveries in the field. Since then, the Forum has undergone a number of changes to reflect how quantum ICT has progressed, and it now comprises more than 100 members, including both individuals and companies, who lead the quantum-related R&D. The Forum hopes to increase its membership, and many of its activities—from its regular public seminars to interviews with leaders in the field—are dedicated to doing so.

The Forum is an ideal organization for people with creative ideas seeking networking opportunities in Japan, looking for research partners, and pursuing funding for their ventures. It also provides an opportunity to have real influence on the direction of quantum ICT. For example, the Forum's input is requested at the task force meeting on the Cross-Ministerial Strategic Innovation Promotion Program, a Japanese government initiative that has identified key technology fields—such as quantum ICT—that have a strong potential to grow the economy.

Tomita stresses that although the Forum is focused currently on Japan, it also provides a place for international partners to understand the talent and dedication that Japan has to offer to quantum ICT.

"It is important for Japan not only to have a strong environment for quantum ICT; we also want the world to know how much investment the country is making in this field and how dedicated we are to it," he says.

Sponsored by

Quantum ICT Forum

Scientific discoveries for technological advances

Throughout its history, Toshiba technology has been found everywhere, ranging from everyday items in the home to social infrastructure to protect entire nations. This diversity reflects not only the high quality of its products, but also the commitment and excellence of its research, which pushes for discovery in both applied technology and pure science. Now, Toshiba's scientific breakthroughs are impacting computer science, setting the stage for faster, more accurate computational solutions to businesses' most difficult problems.

Protecting data

Toshiba is a global company driving innovation in its many research centers across the world, including its Cambridge Research Laboratory in the United Kingdom, founded more than 30 years ago. It was there that Toshiba made its first significant innovation in quantum cryptography.[1] With quantum computing becoming a reality and the protection of data becoming more urgent, Toshiba is leveraging its initial investment in the field to generate estimated revenue exceeding USD 3 billion by 2030.

"Data vulnerability is a major concern in many sectors. Our quantum key distribution [QKD] platform is being used by governments, health care centers, and financial companies to protect data," says Yoshimichi Tanizawa, chief research scientist at Toshiba.

Using its QKD, Toshiba has repeatedly set world records for the fastest data transmission time over long distances, drawing attention from many organizations for data encryption.

In Japan, Toshiba, Tohoku University, and NICT demonstrated a secure genome analysis data backup to multiple sites with Toshiba's QKD system.[2] In the United States, JPMorgan Chase, Toshiba and Ciena demonstrated QKD system securing a peer-to-peer blockchain network.[3] In London, BT and Toshiba, along with EY, launched the trial of world first commercial quantum-secured metro network, helping secure information transmission across the city's data networks.[4] And in South Korea, Toshiba group and KT is developing a long distance hybrid QKD network, enabling the safe transfer of data between the country's biggest cities, Seoul and Busan.[5]

Changing science and technology

While Toshiba research is designed to push new frontiers in technologies, it is also changing our understanding of the science behind these technologies. Such is the case with its simulated bifurcation machine. In their efforts to develop a quantum annealer, which is considered the gold standard for solving combinatorics optimization problems, Toshiba scientists discovered that the principle of quantum annealing called "adiabatic evolution" could be applied to classical nonlinear Hamiltonian systems. The result was the simulated bifurcation algorithm,[6] which incorporates bifurcation phenomena, adiabatic evolution, and chaos theory, allowing digital computers to quickly solve the same complex problems thought exclusive to the domain of quantum computing.

"This is a new idea in computer science. This algorithm can be used by classical computers to achieve high performance from parallel processing," said Toshiba chief research scientist Hayato Goto, who discovered the algorithm. The incorporation of parallel processing distinguishes simulated bifurcation machines from other simulated annealers used to solve optimization problems.

"In simulated annealers, we have to update variables one by one. But with this algorithm we can update the variables in parallel, which is a tremendous advantage," he added. Indeed, his algorithm has set world records for computation times, and problems that would take more than a year to solve using standard central processing unit (CPU)-based simulated annealers can be completed in less than an hour if using the simulated bifurcation algorithm.[7]

Thus, industries anticipating the benefits of quantum computing—including finance, logistics, and pharmaceuticals—no longer need to wait when working with Toshiba.

QKD and the simulated bifurcation algorithm are just two examples of Toshiba's efforts in quantum computing and another reminder of how its research breaks new ground in science and technology. The drive for this progress, however, comes because Toshiba is committed to people and to the future.

References

1. Z. Yuan *et al.*, *Science* **295**, 102–105, https://www.science.org/doi/full/10.1126/science.1066790.
2. https://www.global.toshiba/ww/technology/corporate/rdc/rd/topics/21/2108-02.html
3. https://news.toshiba.com/press-releases/press-release-details/2022/JPMorgan-Chase-Toshiba-and-Ciena-Build-the-First-Quantum-Key-Distribution-Network-Used-to-Secure-Mission-Critical-Blockchain-Application/default.aspx
4. https://www.global.toshiba/ww/news/corporate/2022/04/news-20220427-01.html
5. https://www.global.toshiba/ww/company/digitalsolution/news/2022/0328.html
6. H. Goto, K. Tatsumura, A. R. Dixon, *Sci. Adv.* **5**, eaav2372 (2019), https://doi.org/10.1126/sciadv.aav2372
7. H. Goto *et al.*, *Sci. Adv.* **7**, eabe7953 (2021), https://doi.org/10.1126/sciadv.abe7953

Sponsored by

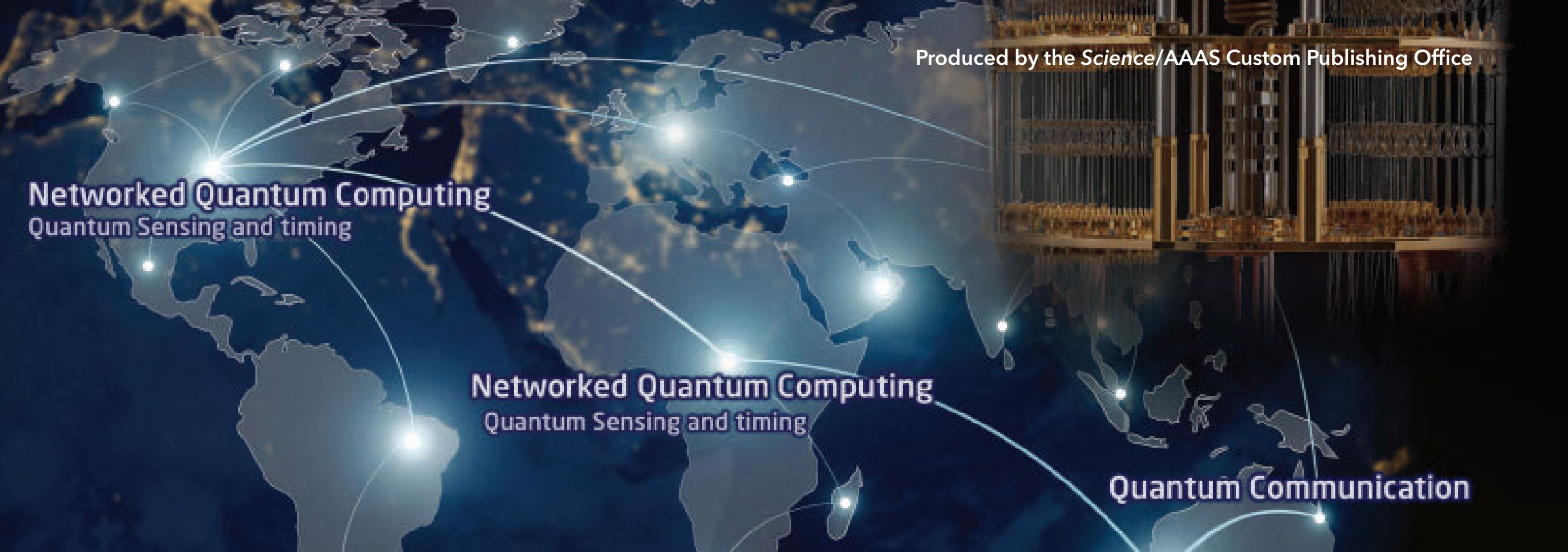
TOSHIBA



Simulated bifurcation machine

PHOTO: PROVIDED BY TOSHIBA

Produced by the *Science*/AAAS Custom Publishing Office



A company of firsts in quantum technology

Throughout its history, NEC—originally founded as the Nippon Electric Company more than 100 years ago—has proudly accomplished a number of firsts in science and engineering. While its initial innovations were in telecommunication technologies, its R&D has positioned the company as a global leader in communications, with developments in semiconductors, mobile communications, and now the next great frontier—quantum technology.

Despite being a multinational company with more than 100,000 employees, NEC conducts research with an academic approach in which curiosity and exploration are paramount. It is this attitude that has led NEC to numerous quantum-computing firsts—most notably, in 1999, being the first in the world to report a solid-state qubit.

The evolution of quantum technology at NEC

Progress in quantum technology is happening in stages. The first is quantum key distribution (QKD), which will revolutionize cryptography, securing data that implements this technology while enabling safety guaranteed by quantum mechanics and information theory.

NEC's quantum cryptography efforts have been dedicated on the commercialization and social implementation of the BB84 system, which excels in medium-to-long-distance communication, and currently center on researching the continuous variable QKD (CV-QKD) system for widespread use, with the aim of further expanding its applications. NEC has been conducting demonstration experiments in various fields, such as health care and finance, with a secure encryption method that will not be deciphered even when quantum computers eventually appear. Many enterprises with a long-term perspective on their business have been interested in NEC's quantum cryptography technology as a powerful measure to protect their critical information in the future.

"As a company with strong research and development, we are committed to innovation and translating that innovation into customer solutions," says Yuichi Nakamura, an NEC executive professional and former vice president.

Another essential step toward quantum technology is to control quantum states with accurate clock timing despite environmental disturbances. This requires quantum clocks that operate accurately against external turbulence. Researching the construction of these clocks, NEC has reported specific coherent-population-trapping resonances.[1] Advances in this technology will have profound effects on the robustness of our communications systems by protecting them against natural disasters and malicious attacks.

The next critical step for quantum technology is quantum computing. In 1999, NEC succeeded in demonstrating the basic operation of the world's first superconducting-based quantum bits suitable for integration, and has continued to research and develop quantum computers for more than 20 years.[2] Currently, we are developing a quantum annealing machine dedicated to combinatorial optimization problems using Josephson parametric oscillators, as well as a fault-tolerant quantum computer that no one has been able to produce yet. Since 2019, we have advanced quantum technology by starting to verify the application of solving customers' combinatorial optimization problems with simulated annealing using vector machines, which are supercomputers that NEC has accumulated over many years.

Great gains through great ambition

NEC's work demonstrates the incremental process required to achieve true quantum technology. Its competence is on par with those who are developing the "quantum internet." However, as with many of the grander problems in science today, no one organization can solve them all. While NEC is at the forefront of the evolution of quantum technology, the company is therefore partnering with many university laboratories and companies, building a wide network of quantum technology experts who are dedicated to realizing this great leap in quantum ability.

This strategy, Nakamura explains, is in place because NEC's ultimate goal goes beyond individual applications such as QKD or quantum computing. "Few companies have the diversity in quantum technology that we have. And few companies have the partners we have," he says. "Ultimately, our technology aims to solve today's most difficult challenges and meet society's most pressing needs. Quantum technology definitely has such potential."

References

1. K. Matsumoto *et al.*, *Phys Rev. A* **105**, 023110 (2022).
2. Y. Nakamura, Y. A. Pashkin, J. S. Tsai, *Nature* **398**, 786–788, https://www.nature.com/articles/19718.

Sponsored by

\Orchestrating a brighter world NEC

PHOTO: PROVIDED BY NEC

PHOTO: PROVIDED BY POSTECH

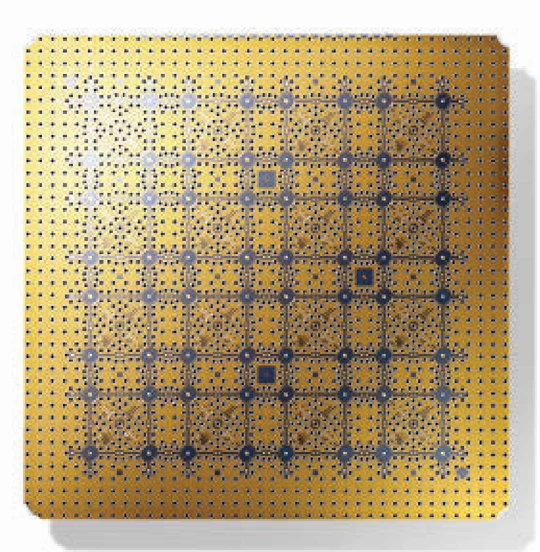
The complete array of advanced computing

Since its founding, the Tokyo-based information company Fujitsu has been using communications technologies to build a more sustainable world. Critical to this aim is staying at the forefront of computing. In the past 20 years, this has meant putting intensive efforts into high-performance computing (HPC) and quantum computing in order to offer the most comprehensive toolkit available for problems demanding computational solutions.

Collaboration is key

Fujitsu recognizes that progress comes faster by working with partners who hold complementary strengths. Its supercomputer Fugaku—which it completed last year in collaboration with RIKEN, Japan's largest research institute—held the top spot in HPC rankings for four consecutive terms. Several groups in Japan are using Fujitsu's supercomputing technologies to interpret tsunami data for evacuation responses and protein bindings for drug design.

In the promising field of quantum computing, Fujitsu has teamed with RIKEN to establish the RIKEN RQC-Fujitsu Collaboration Center, which aims to develop quantum computers using superconducting qubits to solve even more difficult societal problems. Parallel to this, a collaboration with Delft University of Technology in the Netherlands is looking at the advantages of diamond-spin qubits.[1] Where superconducting qubits have excellent scalability so far, diamond-spin qubits can hold quantum information relatively long. A major concern in quantum computing is the sensitivity of qubits to environmental interference, including noise. Fujitsu's collaboration with Delft University also recently revealed a solution to this problem by demonstrating a fault-tolerant qubit operation using a quantum processor based on spin qubits in diamond.[2]



Superconducting 64-qubit chip IMAGE: © RIKEN CENTER FOR QUANTUM COMPUTING

While its projects with RIKEN and Delft are focused on hardware, Fujitsu is teaming with Osaka University in Japan to develop software for fault-tolerant quantum computing in combination with Fujitsu hardware for higher-level quantum computers. The collaboration will take the form of a new research division on the university campus.

While working on superconducting and diamond-spin quantum computers, Fujitsu also developed the world's fastest 36-qubit quantum computer simulator using its supercomputing technologies.

"In quantum computing, there is no best technology. Fujitsu believes it must explore multiple channels to reach the fullest potential," says Shintaro Sato, head of the Quantum Laboratory at Fujitsu.

Quantum-inspired computers

Fujitsu surveys have found that a large majority of businesses are eager to benefit from quantum computing. However, many experts predict that the first commercially practical quantum computers are still years away. As an alternative, researchers are developing quantum-inspired computing technologies. The best example at Fujitsu is its Digital Annealer.

Operational at room temperature and using traditional racks, Fujitsu's Digital Annealer is already being employed in multiple industries, bringing radical solutions and great savings. Successful use cases include new drug design that reduces the delivery of candidate molecules from several months to several weeks, warehouse management that reduces the distances travelled for parts by 45%, and traffic optimization that reduces car congestion by 40%.

Its versatility has also been demonstrated in response to COVID-19. A critical problem in managing the pandemic has been the effective distribution of personal protective equipment, such as masks and N95 respirators, which were suddenly in major demand throughout the world when the pandemic started, leading to huge supply shortfalls. It was validated by a CRADA report that the Digital Annealer optimized resource allocation and minimized the time data traveled to emerging hotspots in the United States in mere seconds, as compared to the days or weeks required by traditional computing algorithms. It was also used to maximize the number of people who could attend professional football matches in Germany while respecting social distance policies, allowing communities to return to normal life faster.

The future of computing

Quantum computing will have profound effects that resonate throughout society. However, as powerful as this technology may one day become, it will not solve all the problems by itself. In other words, quantum computing will not replace HPC, but as a very important new tool, work with other computing technologies to solve difficult societal problems in the future.

"Quantum computing is just one type of computing at which we excel. Our research is motivated by society's needs. In the future, our offerings will include supercomputing, quantum-inspired computing, and quantum computing," says Sato.

References

1. R. Ishihara *et al,* 2021 IEEE International Electron Devices Meeting (IEDM), https://ieeexplore.ieee.org/document/9720552
2. M. H. Abobeih *et al.*, *Nature* **606**, 884–889 (2022), https://www.nature.com/articles/19718.

Sponsored by

FUJITSU

NEWS FROM Science | AAAS



NEWS

IN BRIEF

Edited by
Jeffrey Brainard



Transmission lines carry electricity from a coal-fired power plant in Weisweiler, Germany.

PHOTO: INA FASSBENDER/AFP/GETTY IMAGES

CLIMATE POLICY

Carbon emissions increase—as do ways to track them

Carbon dioxide emissions from burning fossil fuels are on track to rise 1% this year from the 2021 level, making it harder for many nations to reach their goal of achieving net-zero emissions by 2050, scientists from the Global Carbon Project said last week. They cited an easing of pandemic precautions, including increased air travel, as one reason for the rise. Most researchers say the world is unlikely to meet the net-zero goals and limit global warming to 1.5°C by 2050. But two new tools announced last week will aid efforts by improving the ability to track, verify, and regulate greenhouse gases. One tool, developed by the Climate TRACE coalition, uses satellite imagery and machine-learning algorithms to detect and measure emissions from 72,000 sources, including power plants. Separately, the United Nations unveiled the Methane Alert and Response System, which will use data from new satellites capable of spotting large leaks of methane, a powerful greenhouse gas. The announcements came as politicians at the U.N. climate conference in Egypt debated whether and how wealthy countries should pay for climate-related damages to low-income nations.

Wrongful dismissal case is settled

RESEARCH SECURITY | The U.S. government has agreed to pay hydrologist Xiafen "Sherry" Chen $1.8 million to settle a wrongful dismissal lawsuit that stemmed from a failed federal prosecution alleging threats to national security. In 2014, the government accused her of tapping into a restricted database on U.S. water management and providing information to a Chinese government official while working at the National Weather Service. Chen denied acting improperly, and the government dropped the charges in 2015. The Department of Commerce, which includes the weather service, fired her in 2016, but after a review board ruled the termination illegal, put her on administrative leave starting in 2018. Her case mobilized the Chinese American community, which saw it as racial profiling years before former President Donald Trump's administration's China Initiative against Chinese espionage drew similar criticism. In the 10 November settlement, the Department of Commerce does not admit any wrongdoing but agrees to meet with

Chen to discuss how she was treated and to issue a letter praising her work. Chen agreed to retire by the end of this year. Her legal team called the settlement "a great blow … against bigotry and for the rights of Asian Americans."

China trims some COVID-19 rules

PUBLIC HEALTH | China last week announced 20 revisions of pandemic control and prevention measures that somewhat ease the burden on its weary, frustrated population. Changes include cutting required stays in designated quarantine facilities from 7 days to 5 days for international travelers and close contacts of infected people, ending tracing of contacts of patients' contacts, and restricting mass testing to situations where the source of infection is unclear. Local governments retain responsibility for setting the timing, location, and duration of lockdowns, which have disrupted industry and sparked increasingly angry protests; on 14 November, residents of Guangzhou defied a lockdown by crashing barriers and marching through the streets. The policy changes come as COVID-19 is surging again in China: The National Health Commission reported 17,909 new cases on 14 November, the most since the spring. Most of the new cases were asymptomatic.

Roche Alzheimer's drug flops

CLINICAL RESEARCH | An antibody that pharmaceutical giant Roche designed to treat Alzheimer's disease by targeting beta amyloid, a protein that builds up in patients' brains, has failed in two large, phase 3 clinical trials. Compared with a placebo, injections of gantenerumab slowed cognitive decline on standard tests by just 6% or 8% in trials enrolling nearly 2000 people with mild dementia due to Alzheimer's, Roche announced on 13 November. That reduction was not statistically significant. The drug removed less beta amyloid than expected, which some scientists suggest explains its failure. The setback follows positive results earlier this year for an anti-amyloid antibody called lecanemab, made by Biogen and Eisai. More detailed results on several antibody drugs are expected at the Clinical Trials on Alzheimer's Disease meeting later this month.

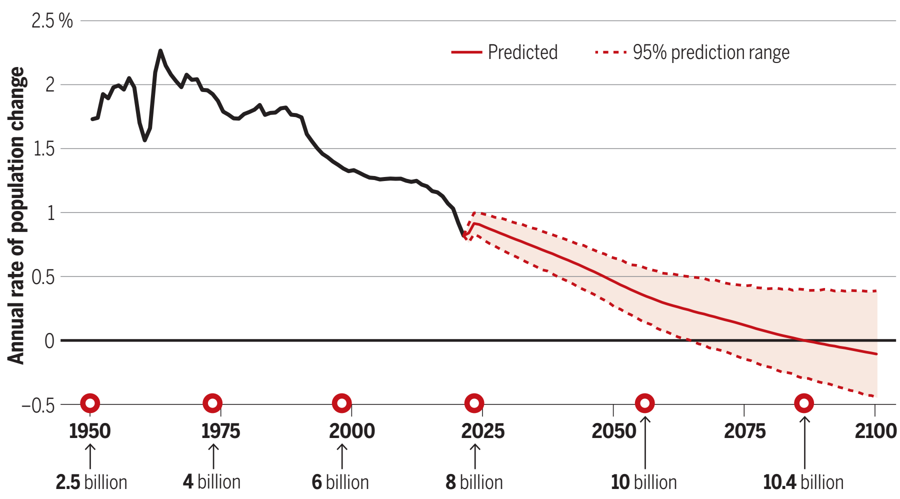
Improved masks win U.S. contest

PANDEMICS | Two small companies that make innovative face masks designed to thwart the spread of pathogens tied for first place this week in a U.S. government–sponsored competition that awarded them each $150,000 to further develop the protective wear. The Mask Innovation Challenge, bankrolled by the Biomedical Advanced Research and Development Authority (BARDA), tested the masks for ability to filter out airborne particles as small as viruses and for breathability, comfort, and looks. The contest began in March 2021 and attracted 1448 entrants. One first-place winner, called Airgami and made by Air99, uses an origami shape that provides a big breathing space, making it comfortable to wear for long periods. The other, ReadiMask, by Global Safety First, has no straps and uses adhesive to make a tight fit on differently shaped faces. Both masks are already on the market. Despite the wins, BARDA says it has no plans to purchase either to stockpile for health emergencies.

DEMOGRAPHY

World population hits 8 billion as growth slows

Earth's population has reached a milestone by surpassing 8 billion people, the United Nations said this week. But the rate of increase is falling, and global population may begin to decline late in the century after topping out at about 10.4 billion, according to the U.N. Population Division. Its *World Population Prospects 2022* report notes that two-thirds of the global population already lives in a country or area where lifetime fertility is below 2.1 births per woman, roughly the level required for zero growth for a population with low mortality. More than half of the projected increase in global population between now and 2050 will be concentrated in just eight countries: the Democratic Republic of the Congo, Egypt, Ethiopia, India, Nigeria, Pakistan, the Philippines, and Tanzania.



THEY SAID IT

> "It just tells us how terrible our culture is becoming, that we can't have an honest scientific debate."

Georges Benjamin, executive director of the American Public Health Association, in MedPage Today, on the decision by public health specialist and commentator Leana Wen not to speak at its annual meeting about harassment of public health officials, after she received criticism and threats over her views about COVID-19.

Big satellite vexes astronomers

SPACE POLICY | Scientists are alarmed that the glow from the largest commercial communications satellite could interfere with ground-based observations. The satellite, BlueWalker 3, unfurled its 64-square-meter antenna this week, which made it among the brightest satellites in the sky. BlueWalker 3 is a prototype for the world's first space-based broadband network, planned by the company AST SpaceMobile, which would deploy a constellation of 168 even larger satellites. Astronomers worry it could blot out objects such as exploding stars or Earth-bound asteroids. Radio astronomers are also troubled because the satellites will operate at radio frequencies that could infringe on parts of the spectrum traditionally reserved for the ground-based observatories. Astronomers were already anxious that communications satellites launched by the corporation SpaceX, which plans a network of thousands, are obstructing observations.

CREDITS: (GRAPHIC) C. BICKEL/*SCIENCE*; (DATA) UNITED NATIONS, DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS, POPULATION DIVISION, *WORLD POPULATION PROSPECTS 2022*



IN DEPTH

U.S. ELECTIONS

Science community braces for divided government

Split control of Congress could produce fiery hearings and limit new funding

By **Jeffrey Mervis**

The U.S. Congress returned this week after tumultuous midterm elections that left the Republicans likely to retake control of the House of Representatives and the Democrats beating expectations by retaining the Senate.

A divided Congress could mean a bumpy ride for the U.S. research community over the next 2 years. In the short run, however, science advocates hope the looming prospect of gridlock could propel legislators to approve bigger research budgets before the current session ends next month. Once the new Congress begins its 2-year session in January 2023, funding increases could become more difficult to secure.

"I believe that the days of generous science funding increases are over," says Robert Atkinson, president of the Information Technology and Innovation Foundation.

The 8 November elections didn't generate a "red wave" that would have given Republicans the sheer numbers to roll back large parts of President Joe Biden's agenda. Instead, neither party will likely be able to advance major new policy initiatives.

House Republicans have said they will aggressively investigate whether the Biden administration is doing enough to keep pace with China's push to become a technological superpower, how the National Institutes of Health (NIH) responded to the COVID-19 pandemic, and whether a laboratory leak in China led to the catastrophe. They will also likely challenge the Biden administration's efforts to curb climate change by moving away from fossil fuels. Although such hearings will allow Republicans to showcase political arguments that resonate with their supporters, they are unlikely to lead to substantive policy changes.

Science advocates are hopeful the partisan battles and gridlock won't undermine the traditional bipartisan support for research funding. An early test began this week as the current Congress took up a massive piece of legislation that would set spending levels for all federal agencies in fiscal year 2023, which began on 1 October. (Federal agencies are now under a spending freeze that expires on 16 December, raising the prospect of a government shutdown unless the freeze is extended or replaced with a yearlong appropriation.)

A new law, the CHIPS and Science Act, calls for double-digit annual funding boosts for several research agencies, including the National Science Foundation (NSF), and science advocates hope lawmakers will follow through on that plan. They'd also like to see NIH's budget keep pace with inflation—or exceed it. Several higher education lobbyists think the Democrats' stronger-than-expected showing at the polls will result in 2023 budgets for research agencies that are close to the generous requests the White House submitted to Congress earlier this year.

Some research advocates fear those numbers could represent a high-water mark, however. Under one scenario, if the Democrat-led Senate and the Republican House fail to agree on new spending levels in 2024 and 2025, budgets could end up essentially frozen at or near the 2023 levels.

"Many Republicans, and some Democrats, view the massive expansion of the national debt in the last 3 years as unsustainable," Atkinson says. Those lawmakers "will likely try to limit the growth of nonentitlement, nondefense spending to at most the rate of inflation."

The fate of future research budgets will be partly in the hands of the lawmakers who end up leading the appropriations panels in both chambers. Senator Patty Murray (D–WA) is expected to succeed retiring Senator Patrick Leahy (D–VT) atop the full Senate spending panel, with subcommittee chairs still in flux. A Republican House is expected to elevate Representative Kay Granger

PHOTO: MICHAEL CIAGLO/GETTY IMAGES

(R–TX) to lead the spending panel, replacing Representative Rosa DeLauro (D–CT). Representative Robert Aderholt (R–AL) is in line to lead the panel overseeing NSF, NASA, and several other research agencies, while Representative Tom Cole (R–OK) would get the appropriations subcommittee overseeing NIH.

Cole has been a reliable advocate for biomedical research spending. And one retired Republican congressman thinks his former colleagues will continue to back hefty budgets for NIH. "Who wants to fight with their constituents when they come to Washington to demand the government do more to find a cure for this or that disease?" says Charlie Dent, who now serves on the board of Research!America, an advocacy group for biomedical research. At the same time, Dent says, the retirement of Senator Roy Blunt (R–MO) means NIH needs a new champion in the Senate.

The House science committee, which helps set policy at major nonhealth research agencies, would also get a new leader: Representative Frank Lucas (OK). Currently the panel's top Republican, Lucas worked closely with the committee's outgoing chair, Representative Eddie Bernice Johnson (D–TX), to craft broadly bipartisan bills. That includes CHIPS, although Lucas reluctantly voted against it after Republican leaders decided to enforce party discipline for political reasons.

The science committee is expected to look at how the Biden administration is implementing several popular provisions in CHIPS, notably programs to spread federal research spending to regions of the country that traditionally receive little of it and to accelerate the commercialization of basic research discoveries, creating new industries and lots of well-paying jobs. Issues important to Lucas's rural district are also high on his agenda, including reauthorization of a major bill governing U.S. agricultural research policy, weather programs, and the regulation of drones.

Given the economic and fiscal struggles facing the nation, U.S. researchers shouldn't expect to get everything they want from the new Congress, says John Culberson, a Texas Republican who chaired the House spending panel that oversees NSF and NASA before losing his House seat as part of a Democratic wave in 2018. But Culberson, now a lobbyist for Federal Science Partners, believes Republicans who are likely to occupy key positions in the next Congress "understand that increased support for basic science and space exploration are good for the economy and important to the nation. And they will fund as much science as the country—and taxpayers—can afford." ■

CANCER RESEARCH

Tumors can teem with microbes. But what are they doing there?

New study suggests microbiomes can promote cancer by suppressing immune response and seeding metastases

By **Gunjan Sinha**

Our bodies harbor countless microbes—and so do our tumors, it turns out. Over the past 5 years, researchers have shown cancer tissue contains entire communities of bacteria and fungi. Now, it appears some of the bacteria may be cancer's accomplices.

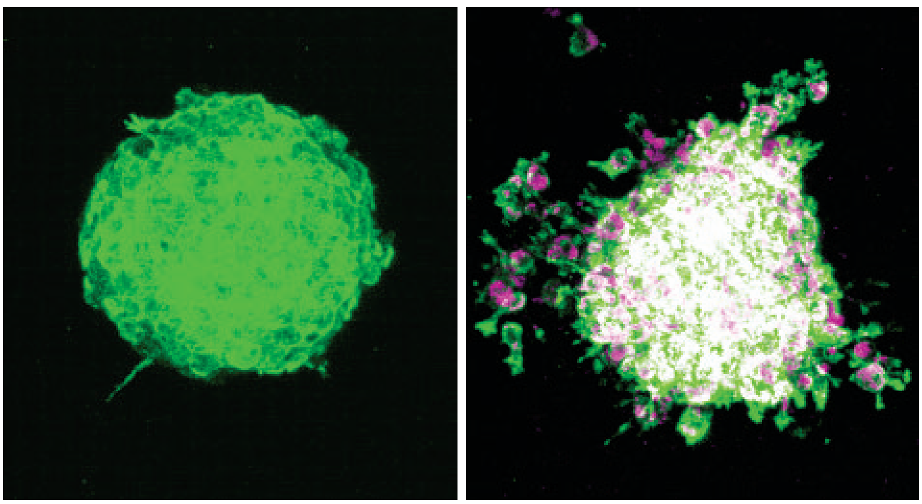
In a paper in *Nature* this week, a team led by Susan Bullman of the Fred Hutchinson Cancer Center reports that in oral and colorectal tumors, bacteria live inside cancer cells and boost their production of proteins known to suppress immune responses. The microbial interlopers may set off a chain reaction that prevents the immune system from killing cancerous cells, and they may also help cancer metastasize to other parts of the body.

The study doesn't entirely clinch the case for a bacterial role in cancer, but it is very suggestive, says Laurence Zitvogel, a tumor immunologist at the Gustave Roussy Institute. "It shows that bacteria in colorectal and oral tumors can actively disturb the immune equilibrium," she says.

Confirmation that microbes can cause tumors to grow or spread could open up new ways to make cancer treatment more effective, for instance by killing bacteria with antibiotics. And because each type of cancer appears to come with a unique microbiome, researchers are exploring whether microbes could be used as a diagnostic tool to detect cancer early in a blood sample.

Until recently, most cancer researchers believed tumors were sterile, says Ravid Straussman, a cancer researcher at the Weizmann Institute of Science. But about a decade ago, as a postdoc at the Broad Institute, Straussman accidentally discovered that human pancreatic and colorectal cancer cells grown in the lab stopped responding to a cancer drug named gemcitabine when *Mycoplasma* bacteria were present in the culture. The bacteria, he discovered, "protected" the cells by producing an enzyme that breaks down gemcitabine.

Straussman found he could render gemcitabine ineffective in mice with colon cancer by injecting the animals with other types of bacteria, including an *Escherichia coli* strain, and that treating them with antibiotics restored the drug's effectiveness. When he studied 113 human pancreatic cancer samples, he found bacteria that produced the drug-chewing enzymes in 76% of them—raising the question of whether they contributed to drug resistance in human cancers. Straussman and



Researchers cultured cancer spheroids without bacteria (left) and with *Fusobacterium nucleatum* (shown in pink, right). After 20 hours, individual cancer cells containing bacteria migrated away from the spheroid.

IMAGE: JORGE GALEANO NINO/BULLMAN LAB

his colleagues are now planning a clinical trial to test whether antibiotics can improve pancreatic cancer treatment.

Soon afterward, Gregory Sepich-Poore, a doctoral student in the lab of microbiome researcher Rob Knight at the University of California, San Diego (UCSD), was hunting for ways to diagnose pancreatic cancers early. He was motivated by his grandmother's death from the cancer, which is often diagnosed too late for treatment to be effective. Inspired by Straussman's 2017 paper, Sepich-Poore began to scour the Cancer Genome Atlas, a large DNA database of human cancers, for snippets of genetic material from microbes.

In March 2020, he, Knight, and colleagues reported that microbial RNA and DNA was present in each of the 33 types of cancers they studied, and that each cancer type had a unique microbiome. The team also found those distinct microbial signatures in blood samples from cancer patients. Based on their findings, Sepich-Poore and Knight co-founded San Diego–based Micronoma, a startup that aims to identify early-stage cancer in blood samples—a so-called liquid biopsy.

Later in 2020, Straussman and his colleagues confirmed that many tumors have distinctive populations of microbes and found they mostly reside inside cancer and immune cells, rather than between those cells.

Fungi often take up residence in tumors as well. In a study of 17,000 tumors, published in *Cell* in September, the UCSD and Weizmann groups found fungal species residing in each of 35 cancer types. Again, each cancer type was associated with a distinct combination of species, which could help refine Micronoma's diagnostic tools. (Straussman now sits on the company's scientific advisory board.)

The paper reported another striking finding: Certain combinations of fungal species correlated with lower odds of survival in several types of cancers, most strongly in ovarian and breast cancer. In October, another group reported something similar in *Cancer Cell*: The presence of a particular bacterial signature seemed to hasten death in pancreatic cancer. The probability of surviving 2 years after treatment doubled in patients that did not have the signature. "That's an eyebrow-raising finding," says co-author Martin Blaser, a cancer microbiome researcher at Rutgers University, Piscataway, who also sits on Micronoma's scientific advisory board.

But none of these findings showed just how fungi or bacteria might lead to a worse outcome. Now, Bullman and her colleagues have addressed the question by studying eight tumors removed from patients with oral cancer and 19 others from colorectal cancer patients. Mapping the distribution of the microbes showed they only colonized specific areas of the tumors. These infected regions had high levels of proteins known to suppress cancer-fighting T cells or fuel cancer growth. T cells amassed outside these regions, the researchers found, but few were found inside. (Instead, the regions contained neutrophils—a type of immune cell that fights infections, among other jobs.) "It's conceivable that the bacteria are somehow causing the T cells to move away from the tumor," Blaser says.

Using a technique called single cell sequencing, the researchers found bacteria preferentially infect cancer epithelial cells—which line the inside surface of organs—and that only cells in which *Fusobacterium* and *Treponema* bacteria were dominant tended to show both immunosuppressive and cancer promoting characteristics.

> **"This paper has taken the field a big step forward."**
> **Ravid Straussman**, Weizmann Institute of Science

"This paper fills a critical gap" by showing that bacteria inside cancer cells may alter the cells' behavior, says George Miller, a cancer doctor and researcher at Trinity Health of New England.

Bullman and her colleagues also co-cultured *Fusobacterium* species with colon cancer spheroids—small models of human cancers—embedded in a matrix that contained neutrophils, and compared them with bacteria-free spheroids. With the bacteria present, neutrophils tended to move toward the cancer cells, just as they did in the patient tumor samples. And the researchers saw infected cancer cells breaking off the spheroids and migrating, which Bullman thinks may be a sign that they are metastasizing.

Zitvogel says the paper paints a plausible picture of how microbes could hamper the body's defenses against cancer. Still, the spheroid model "is a reductionist approach," she cautions; the human body, which has a varied arsenal of immune cells and a diverse and largely beneficial microbiome, may have other mechanisms that keep cancers from metastasizing.

The study was small and only included two types of cancers, Straussman adds, which leaves plenty of work to do. But, "Bullman's research has shown us how we should be exploring the tumor microbiome," he says. "This paper has taken the field a big step forward." ■

Gunjan Sinha is a science journalist in Berlin.

SCIENTIFIC INTEGRITY

Whistleblower finds possible misconduct in his own papers

Matthew Schrag confronts a mentor after their joint work is flagged on the PubPeer website

By **Charles Piller**

When Vanderbilt University neuroscientist and physician Matthew Schrag went public earlier this year with concerns about apparently doctored images in scores of Alzheimer's papers—including seminal research underpinning one aspect of the dominant amyloid hypothesis of the disease—he anticipated that his motives and analyses would be dissected. "I also expected every project that I ever participated in to be carefully scrutinized, and that my work would stand up to that scrutiny," he says.

So Schrag assumed there would be an innocent explanation when, a few weeks after a *Science* investigation reported his disturbing findings of apparent misconduct, he received automated emails from PubPeer, a web forum where scientific wrongdoing charges are often leveled. They notified him that two of his own articles from more than 15 years ago had been flagged as containing dubious images.

On close examination, Schrag had to confront an unnerving prospect: that a co-author, neuropharmacologist Othman Ghribi, Schrag's first mentor and still a trusted friend, might have also engaged in misconduct. The papers, published in 2006 when Schrag was an undergraduate working in Ghribi's lab at the University of North Dakota (UND), covered research on several factors related to amyloid proteins in rabbit brains. (Many Alzheimer's researchers believe the disease is caused by amyloid's effects on brain cells.)

Schrag soon discovered that the suspect work in the two papers fit a large pattern of questionable research spanning much of Ghribi's career, both before and after they worked together.

According to Schrag, in a phone conversation the senior scientist emotionally acknowledged "problems" in many of his papers, including those two flagged on PubPeer, and accepted responsibility. During their discussion, which Schrag recounted to *Science*, Ghribi maintained to his former student that the underlying findings were correct but admitted to exaggerating data. "I'm nauseated talking about it," Schrag says.

After initially agreeing to an in-person interview with *Science*, Ghribi backed out a few days before the appointment, citing a pending UND investigation. He did not respond to a request to verify Schrag's account of their talk or to comment on a dossier of suspect images in his papers that Schrag has compiled with other forensic experts. But Ghribi told *Science* via a 5 October email that he wanted to "exonerate [Schrag] of any wrongdoing with the manuscripts he co-authored in my lab."

Back in August, Schrag contacted a UND official to see whether any of the original images for the two suspect papers could be found. The official said he could not identify any relevant documents. But the university found that Schrag's concerns about the papers, along with PubPeer comments on those and other publications by Ghribi, warranted the inquiry, according to a 10 October email to Ghribi, obtained by *Science*.

"We are taking all reasonable steps to secure records related to the research in question," John Mihelich, a UND vice president, told *Science* in an email. "We are in communication with the appropriate federal research offices and sponsors."

Resolving the matter might prove challenging. Ghribi has moved to the University of Texas (UT), Rio Grande Valley. And Schrag says during the recent phone call Ghribi said that, out of remorse, he had discarded his scholarly awards and purged his computers and files of raw experimental data. If so, a review of Ghribi's papers could be hampered, because proof of image manipulation sometimes requires uncropped original images for comparison with published images.

Schrag recently asked two journals to retract his suspect papers with Ghribi. The publisher of *Experimental Neurology* said via email he was awaiting comment from Ghribi. Andrew Lawrence and Marco Prado, editors of the *Journal of Neurochemistry*, said in an email to *Science* they "applied forensic analysis and concluded there are issues with the publication." They added that all authors, including Ghribi, agreed the paper should be retracted, which will occur soon. Schrag says he's also evaluating whether images


Matthew Schrag (left) and Othman Ghribi enjoying better times at Schrag's Ph.D. graduation ceremony.

PHOTO: MATTHEW SCHRAG

prepared by Ghribi for a third joint paper, which appeared in 2008 in the journal *Hippocampus*, also warrant a retraction or correction; he has contacted the journal to share his concerns.

After reviewing additional comments posted to PubPeer about other Ghribi papers, Schrag says he recently resolved to take a deeper look at his mentor's work, including all their joint papers. He enlisted help from microbiologist and forensic image analyst Elisabeth Bik and another image sleuth—a nonscientist who uses the pseudonym Cheshire, in part to minimize legal risks. (*Science* confirmed Cheshire's identity and agreed to keep it confidential.) In recent years, Bik and Cheshire have identified thousands of apparently manipulated or duplicated images in many papers, often leading to retractions or corrections.

In their assessment of Ghribi's work from 2001 through 2019, the three found suspect images in 33 papers, including the three co-authored with Schrag. Ghribi was the only author shared by all the papers, and he was usually in a prominent position as either first or last author. The problematic images included Western blots (a common method for displaying proteins in tissue samples) and micrographs of brain tissue. Their joint 67-page dossier, which Schrag provided to *Science*, shows more than 100 apparently problematic images, many from work funded by grants from the National Institutes of Health. (The agency declined to comment to *Science* on the matter, but in an email to Schrag it said it would look into the concerns.)

The three found possible scientific misconduct in five other Ghribi papers in the *Journal of Neurochemistry*. Lawrence and Prado said those cases "will be also carefully evaluated in a timely way." Four papers with suspect images, co-authored by Ghribi, appeared in the *Journal of Alzheimer's Disease*. UT San Antonio professor George Perry, chief editor of the journal, reviewed the full dossier and says he was surprised to find many images "suspicious," although he emphasizes that he has no expertise in forensic image analysis. He describes Ghribi as a "likeable," relatively established researcher whom he has always considered "a person of integrity."

Perry says he contacted Ghribi to ask for more information on the suspect papers and discuss Ghribi's role as a member of the journal's editorial board. The journal will assess the claims, he adds, and issue errata or retractions if warranted.

Schrag acknowledges Ghribi's early influence in his career and feels sad about his former mentor's apparent misdeeds. But Schrag says he faced an imperative to correct the scientific record, including work he played a part in. "You have to have a near-religious commitment to research integrity. If the rules apply to others they have to apply to [all of] us," he says. ■

This story was supported by the *Science* Fund for Investigative Reporting.



Woolly mammoth tusks, such as these from Wrangel Island in Russia, may be sustaining the global market for illegal elephant ivory.

PHOTO: SERGEY GORSHKOV/MINDEN PICTURES

CONSERVATION

Booming trade in mammoth ivory may be bad news for elephants

Paleontologists are urged to take a stand against a market that may provide cover for continued poaching

By **Michael Price**, *in Toronto*

In 2015, Andy Huynh was accompanying wildlife guards in Kenya's Maasai Mara National Reserve to help ward off poachers. Fresh off a decade of service in the Middle East with U.S. Special Operations Forces, he thought there was little that could faze him. But when he saw his first poached rhinoceros, with half of its face sawed away for the horn, he turned and threw up. "I knew then and there I wanted to dedicate my life to stopping wildlife crime," Huynh said.

He began to work with various wildlife protection nonprofits, then joined a series of U.N. and Interpol undercover operations in China and Vietnam to bust up the illegal trade in elephant ivory. Now, he has extended his definition of wildlife to the distant past: the great, tusked mammoths and mastodons of the ice age.

At the annual meeting of the Society of Vertebrate Paleontology (SVP) here last week, Huynh argued that the growing trade in ivory from the ancient carcasses found in thawing Arctic permafrost is sustaining a global market that leads to the death of living elephants. He urged paleontologists to raise their voices against the fossil ivory trade—and avoid dealing with unscrupulous collectors who might be involved in it.

Some researchers question whether there are enough data to prove ancient ivory really is buoying the demand for elephant tusks, but others at the meeting welcomed his call to action. "Andy did a great job of making his case," says Thomas Holtz, a paleontologist at the University of Maryland, College Park. "Plenty of SVP talks address mass extinction or even the death of an individual, but those are separated from us by an immensity of time. This dealt with death and suffering happening right now."

Ivory from African and Asian elephants commands up to $3000 per kilogram on the black market, primarily in Southeast Asia, where it is mixed into traditional medicine as a powder and carved into status-signifying statues and other trinkets. About 55 African elephants are killed every day for their tusks. But stricter poaching laws and China's closing of legal ivory carving facilities in 2018 have made it harder for suppliers—sometimes tied to criminal syndicates—to source elephant ivory, Huynh said. "In order to make up for decreasing supply, organized crime has turned to using mammoth ivory."

According to Huynh, criminal organizations in Russia pay good money for private tusk hunters to find and extract mammoth ivory from the melting permafrost. "The rest of the skeleton is either destroyed on the spot or lost to erosion," he said.

He noted a 2014 report from wildlife conservation researchers Lucy Vigne of Oxford Brookes University and Esmond Bradley Martin, which found that since 2002, sales of mammoth ivory had grown from almost nothing to about 40% of all ivory items sold in Beijing and nearly 70% in Shanghai. Huynh also presented more recent, unpublished data from a 2018–19 joint operation conducted by Interpol and the United Nations Office on Drug and Crime that suggested both elephant and mammoth ivory are continuing to enter Vietnamese and Chinese markets, primarily on Russian shipping containers that at times also contain illicit drugs and weapons.

The Convention on International Trade in Endangered Species of Wild Fauna and Flora monitors and regulates the trade of items such as elephant ivory and rhinoceros horn, but it has no explicit protections for extinct animals such as mammoths or woolly rhinos. "So, the trade for mammoth ivory is left unchecked and growing … and the global criminal network has clearly taken advantage of this loophole," Huynh said.

The harm extends to today's elephants, he argued. He noted that although mammoth ivory has surged into the market, the overall demand for elephant ivory appears to have remained constant or even grown. His conclusion: Ancient ivory isn't replacing elephant ivory, but is instead sustaining the market's appetite for it. Some traffickers may even be passing off elephant ivory as legal mammoth ivory, he says, noting that it can be hard to distinguish between the materials in smaller worked pieces, such as trinkets and beads.

Vigne, who did not attend the conference but watched a recording of Huynh's talk, says it's unclear just how mammoth ivory is affecting demand for elephant ivory in mainland China. Some research hints that the rise in mammoth ivory has led to less elephant poaching. Vigne's own visits to mainland Chinese ivory markets suggest "mammoth ivory has certainly helped reduce elephant ivory, but it has also provided a route to [continue to sell] elephant ivory," she says. "So, it's a tricky one."

What is needed, Vigne says, is more information on just how much worked elephant ivory is being trafficked under the guise of legal mammoth ivory. Victoria Herridge, a paleontologist at the Natural History Museum in London, agrees. "You need passionate, activist voices [like Huynh's], but you also need data."

Thomas Carr, a paleontologist at Carthage College and Huynh's undergraduate mentor, notes that paleontology also has a stake in limiting the trade: The sale of fossil ivory reduces the number of mammoth carcasses available for study. "If fossils are being destroyed for tusks, you lose so much data," he says. "It's a loss for science and it's a loss for society."

SVP and the broader paleontological community can help, Huynh says, by issuing public statements and putting pressure on elected officials and international regulating bodies. Scientists should also avoid obtaining samples from unscrupulous tusk hunters, he says, as they might be collaborating with collectors who are employed by criminal organizations. Holtz agrees paleontologists should do more, saying, "I don't think most of us understood that fossils were being run in the same shipments as, for example, heroin."

Jessica Theodor, a paleontologist at the University of Calgary and SVP's outgoing president, says SVP has for decades issued statements decrying the loss of scientifically significant fossils to commercial trade. In the wake of Huynh's talk, though, she says the society will establish a task force to investigate what more it can do to protect mammoth ivory. ■

PHILANTHROPY

Billionaire's crypto company collapse strands scientists

As FTX files for bankruptcy, the grantees its foundations supported may not see all of their pledged money

By **Robert F. Service**

Last week's collapse of the cryptocurrency exchange FTX is sending aftershocks through the scientific community. An undergraduate physics major at the Massachusetts Institute of Technology (MIT) who founded FTX and quickly became a billionaire, 30-year-old Sam Bankman-Fried backed philanthropic organizations that supported a wide variety of science-related causes designed to improve human well-being.

Now, with FTX in bankruptcy and under investigation for misuse of investors' money, his formerly flush foundations are suddenly strapped for cash and much of that work is at risk.

One foundation, the Future Fund, was just launched in February. But by the end of June, its officials reported awarding 262 grants and "investments" totaling $132 million. It's unclear how much of that money has been distributed. But on 10 November, five senior Future Fund officials resigned and announced in a statement, "We are devastated to say that it looks likely that there are many committed grants that the Future Fund will be unable to honor."

"It's definitely a mess," says Josh Morrison, who heads 1Day Sooner, a pandemic preparedness research and advocacy organization that received $375,000 from the Future Fund and the FTX Foundation. During the pandemic, 1Day Sooner became known for advocating so-called human challenge trials, which deliberately infected volunteers with SARS-CoV-2 to test vaccines.

Other notable science recipients of the Future Fund's money include Sherlock Biosciences, which was awarded $2 million for CRISPR-based infectious disease diagnostics; HelixNano, which was awarded $10 million for research on a vaccine effective against all different coronaviruses; and SecureBio, which was given $1.2 million to develop better pandemic defenses, such as an early warning system that screens wastewater for pathogen genetic material.

SecureBio's co-founder, Kevin Esvelt, a biologist at MIT, says the nonprofit is avoiding use of its Future Fund grant money for now, except to pay the salaries of the three newly hired people.

"We don't think it is right that anyone should lose their jobs over a financial calamity totally unrelated to the excellent work they are doing," Esvelt says.

SecureBio is now scrambling to secure emergency funding. The organization just last week released a white paper that reviews new technologies, such as improved personal protective equipment and germicidal lights, that could cope with or halt pandemics. "The events of last week put this critical work in jeopardy," Esvelt says.

FTX's collapse was unthinkable just days earlier. The company, which serves as an online trading platform for cryptocurrency, had assets between $10 billion and $50 billion, according to bankruptcy documents. But it was brought to its knees by an old-style run on the bank, as investors tried to withdraw their money after doubts were raised about FTX's financial health. The company declared bankruptcy on 11 November, and just hours later more than $500 million was reportedly stolen from it by hackers.

Just what will happen to awards the Future Fund and the similar FTX Foundation have already made remains unclear. FTX owes billions of dollars to creditors and is now being investigated by the U.S. Securities and Exchange Commission and the Department of Justice, according to *The Wall Street Journal*.

Writing in an online forum hosted by the Center for Effective Altruism, to which the Future Fund pledged nearly $14 million, Molly Kovite, legal operations manager for the Open Philanthropy foundation, noted that FTX's creditors could try to "claw back" their investments during bankruptcy proceedings. If grantees received awards after 11 August, which is 90 days prior to the bankruptcy filing, "the bankruptcy process will probably ask you, at some point, to pay all or part of that money back," she predicts.

That has grantees wondering how they will pay the bills. "Everyone is obviously really worried," Morrison says. ■



SHELTER FROM THE STORM

A plan to wall off Houston and nearby industry from flooding caused by hurricanes will cost tens of billions of dollars. Will it be enough?

By **Warren Cornwall**, *in Galveston, Texas*

Plans for one of the world's biggest and most expensive flood barriers were born in a second-floor apartment here in this city on the Gulf of Mexico, as water 4 meters deep filled the street below. In September 2008, Bill Merrell, an oceanographer at Texas A&M University, Galveston, was trapped with his wife, daughter, grandson, and "two annoying chihuahuas" in the historic building he owns. Outside, 180-kilometer-per-hour winds generated by Hurricane Ike rattled windows and drove water from the gulf and Galveston Bay into the city.

As saltwater swirled through the shops and restaurants downstairs, Merrell sat in his office and sketched plans for a project he hoped would put an end to the storm-driven flooding that had repeatedly devastated this part of Texas.

It was an ambitious vision: Seventy kilometers of seawalls rising 5 meters above sea level would stretch the length of Galveston Island and beyond. Enormous gates would span the 3-kilometer-wide channel through which ships pass in and out of Galveston Bay. The defensive perimeter would seal off not just Galveston, but the whole bay, with Houston at its far end, protecting more than 6 million people and the country's largest collection of chemical plants and oil refineries.

Though Merrell had spent decades studying ocean currents and storm surges, he had no engineering experience. But as he watched the murky waters soak the city, including his own carefully restored 19th century landmark, he decided there had to be a better way. "The Dutch would never put up with this," he said to his wife.

In 2008, winds and flooding from Hurricane Ike leveled shorefront homes along the Gulf of Mexico in Texas.

Today, that first brainstorm has morphed into a $31 billion plan from the U.S. Army Corps of Engineers, the nation's builder of mammoth water infrastructure. The state of Texas has embraced the idea, creating a taxing district to help pay its share. In July, Congress authorized the Corps to proceed—though it has yet to appropriate money for construction.

The project would be the costliest ever built by the agency. By some measures it would dwarf anything else in the world. The sea gates meant to block gulf waters from Galveston Bay would span a gap bigger than the famous pivoting Maeslant barriers that hold back the North Sea near Rotterdam, Netherlands. "Everything is bigger in Texas," quips Bas Jonkman, a civil engineer and water-control expert at the Delft University of Technology (TU Delft).



Oceanographer Bill Merrell, next to a statue memorializing a deadly hurricane that hit Galveston, Texas, in 1900, developed a plan for protecting the region from storms.

Dutch experts like Jonkman are in high demand these days. Around the world, from New York City to Singapore, governments are planning massive seawalls and other measures to ward off the rising seas and intensifying storms expected from climate change. "A lot of countries are really thinking through now: 'How should we defend ourselves?'" says Marc Walraven, a senior adviser on storm surge barriers to the Dutch government. In the process, each country or city is confronting similar trade-offs: between protecting people and preserving ecosystems, between technically sensible designs and aesthetically acceptable ones, between maximal protection and what's affordable.

In Galveston, Merrell and some other scientists think the Corps hasn't struck the right balance. Merrell warns that its plan—a scaled-down version of his original blueprint—is destined to fail, perhaps catastrophically. "It's too weak—[the defenses] would only stand up to like a 30-year storm," he says. "Essentially you don't have any protection" against the more extreme storms that have already left deep scars on Galveston—and are likely, as climate change advances, to leave more.

LIKE A GYMNAST on a balance beam, Galveston perches on a slender ridge of sand, precarious and exposed. To the north, behind that barrier island, lies Galveston Bay, an estuary half the size of Rhode Island, teeming with shrimp and birds. The bay is so shallow, locals joke that if you fall out of a boat, just stand up. To the south, Galveston faces the Gulf of Mexico, whose warm waters fuel hurricanes nearly every year.

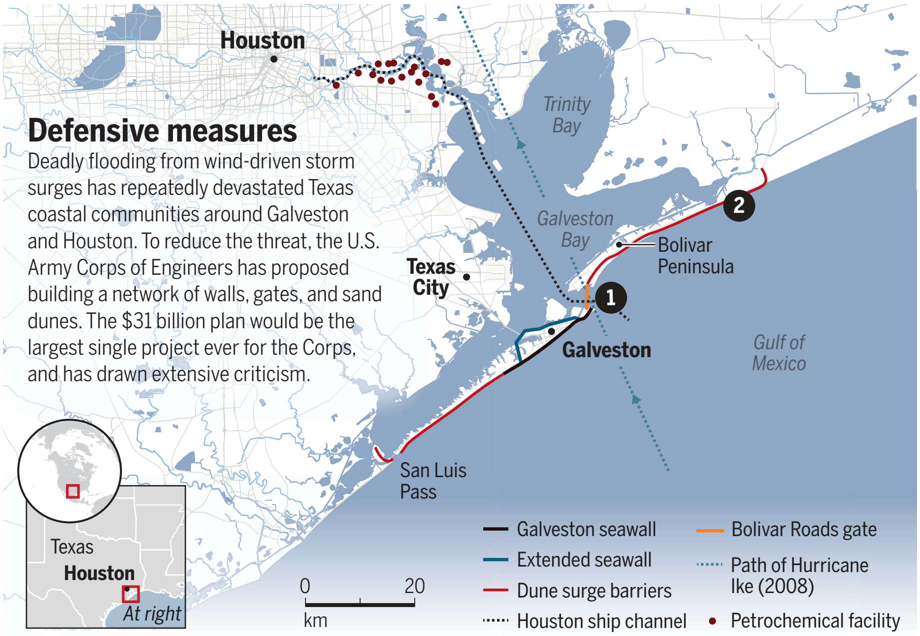
Merrell's 1870 building in the Strand Historic District has survived a number of them, including the all-time worst. In September 1900, a monster Category 4 hurricane with gusts topping 210 kilometers per hour blasted the area. Driven by the wind, flood waters nearly 5 meters deep surged into the city from both the bay and the gulf, crushing thousands of buildings. More than 8000 people died. It's still the deadliest natural disaster in U.S. history.

In the aftermath, local leaders and the federal government erected a 6-kilometer-long, 5-meter-high seawall along the gulf, filling the space behind it with a deep sand layer that sloped gently toward the bay. On that raised ground, they rebuilt their city.

PHOTOS: MELISSA PHILLIP/© *HOUSTON CHRONICLE*; (OPPOSITE PAGE) SMILEY N. POOL/AP PHOTO

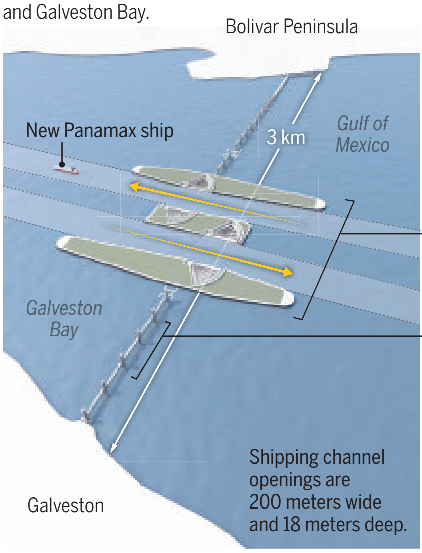
Defensive measures

Deadly flooding from wind-driven storm surges has repeatedly devastated Texas coastal communities around Galveston and Houston. To reduce the threat, the U.S. Army Corps of Engineers has proposed building a network of walls, gates, and sand dunes. The $31 billion plan would be the largest single project ever for the Corps, and has drawn extensive criticism.



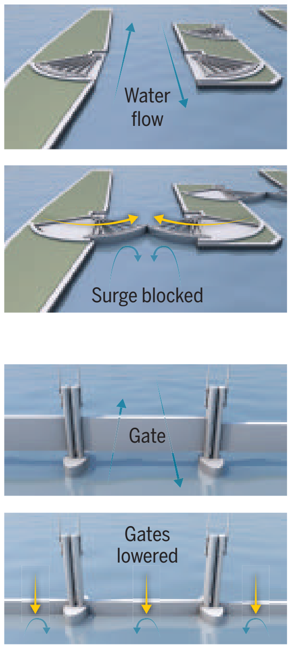
1 Bolivar Roads gate

The centerpiece of the project, a system of gates, would span the 3-kilometer-wide channel that is the main connection between the Gulf of Mexico and Galveston Bay.



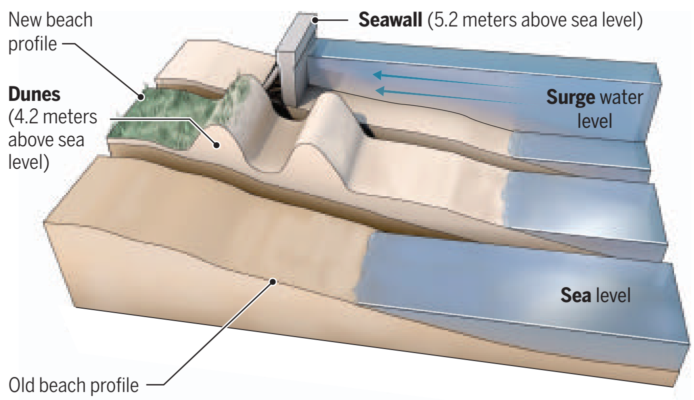
Flood sector gates
To prevent storm-driven waters from surging into the bay, huge gates would close channels that allow freighters and tankers to reach Houston's port and many chemical plants.

Vertical lift gates
Barriers flanking the channels would remain open except during a storm, to allow regular tidal currents to flow back and forth.



2 Dune system

The Corps has proposed protecting the coastline by constructing a double row of dunes with tops up to 4.25 meters above sea level. Some critics fear the dunes won't be tall or strong enough to withstand large storm surges, and would prefer a taller concrete seawall or taller dunes built over a hardened rock core.



GRAPHIC: C. BICKEL/*SCIENCE*

When Hurricane Ike made landfall a century later, Galveston's seawall held. But bay waters flooded central Galveston, including the Strand, from the less protected northern side. Low-lying neighborhoods nearby washed away. Storm surge inside Galveston Bay drenched towns dozens of kilometers from the gulf. When the water receded, it left behind $37.5 billion in damage and 74 dead.

Galveston got lucky. At the last minute, the storm veered east, sending the eye over the city and up the bay. In the Northern Hemisphere, where cyclones rotate counterclockwise, the most destructive winds blow on the eastern side, known as the "dirty side." Sam Brody, a coastal planner at Texas A&M Galveston, calls Ike a "near miss."

A MONTH AFTER the storm, in a small online newsletter, Merrell wrote a column unveiling his idea for a wall that would barricade the barrier islands and Galveston Bay during a storm. He called it the "Ike Dike."

Although the name caught on, the idea didn't at first. Some derided it as a monumental overreach in a place that had relied for years on more modest seawalls to protect towns like Galveston, and on perching houses on stilts. But eventually the Corps, which has a regional office on the outskirts of Galveston, agreed to study the matter. In 2021, it signed off on the current plan, which Congress authorized this summer.

The blueprint has key features in common with the original Ike Dike. At Bolivar Roads, the chief waterway joining the gulf and the bay, four swinging gates, each more than 100 meters long, would guard two openings big enough to fit some of the world's largest freighters, including Panamax ships (see graphic, left). In Galveston, the existing seawall would be extended and improved to encircle much of the city. Overall, analysts at the Corps forecast that the system could reduce storm surge damage to the region by $62 billion over 50 years, or twice its estimated cost.

But the plan's flood-stopping powers are less than what the Corps—and Merrell—first envisioned. In 2018, the agency had proposed a more conventional concrete seawall along the beaches and highways at the gulf's edge, the length of Galveston Island and the Bolivar Peninsula, much like Merrell's Ike Dike.

That proposal did not fare well with island residents. People living near the beaches objected that the wall would be an eyesore, even as it left many homes unprotected. "The public went ballistic," says Kelly Burks-Copes, an ecologist who led the crafting of the plan. "We had some really harsh, really confrontational public meetings."

Recently, standing on a 50-meter-wide beach in front of the houses that crowd

Galveston Island's gulf shore, Burks-Copes explained how the Corps adapted to the criticisms. In its final plan, the agency replaced much of the seawall with two parallel dunes built from sand dredged offshore, each roughly 2.5 meters tall. One dune would crest at 3.7 meters above sea level, and a second, farther up the sloping beach, at 4.25 meters—almost 1 meter shorter than the earlier proposed wall.

The new plan didn't trigger an outcry from locals, but it also doesn't offer them as much protection. The dunes are sized to withstand the kind of storm that, in the historic record, has come along every 50 years on average. The surge from a storm like Ike (which an agency engineer pegged at about an 80-year storm) would flood over the embankments, the Corps acknowledges.

tropical cyclones. But there is broad agreement that the most severe storms—such as Category 4 and 5 hurricanes—will become more intense, says Gabriel Vecchi, a Princeton University climate scientist. And an increasing number of storms in the Gulf of Mexico will be Category 3 or bigger, predicted a 2017 study by scientists at the National Center for Atmospheric Research.

The Corps has understated the risks posed by such storms, Merrell and his collaborators warn, because it overstates the protection the new dunes will offer. "I know from my own experience that those dunes will be completely obliterated by any significant hurricane," says Bruce Ebersole, a civil engineer who until 2011 oversaw storm and flood protection research at the agency's laboratory in Vicksburg, Mississippi.



After Hurricane Ike, officials moved quickly to shore up dunes and seawalls damaged by the storm.

PHOTO: NICK DE LA TORRE/*HOUSTON CHRONICLE*/AP PHOTO

But Burks-Copes said the agency will be able to blunt some of the flooding by closing the Bolivar Road gates at low tide before a big storm arrives, leaving room in the estuary for some of the overflow. Most beachfront homes, meanwhile, are now less vulnerable than when Ike arrived, Burks-Copes noted, because they've been elevated on pilings. "Unless the surge is over the first floor," she said, "they're not impacted."

EVEN UNDER ITS PLAN, the Corps estimates that damages from multiple storms over 50 years could still reach $30 billion or more in the Galveston Bay region. The agency's estimate of how often a storm of a certain size is likely to strike is based, however, on historical patterns. It does not take into account how a warmer future might alter that equation.

Debate and uncertainty surround the many ways a warmer planet could influence

Ebersole, who recently retired from a research position at Jackson State University, worked with Merrell to craft the Ike Dike plan and to vet the proposal from the Corps. He says agency modelers assume the dunes will act like a solid 3.5-meter wall during a storm. But in a major hurricane, "They're going to be eroded and destroyed quite early," he says. "Their effect on storm surge," he suspects, will be "negligible."

To stand up to a 100-year storm, colleagues calculated the dunes would need to be far more massive: nearly 7 meters tall with a crest 45 meters across, almost as wide as a U.S. football field. A more realistic solution, Ebersole and Merrell argue, would be building the original Ike Dike seawall, or an artificial dune made of sand spread over a hardened core of rocks or concrete.

Their view has won support from some important local political leaders, such as Michel Bechtel, mayor of Morgan's Point, a town on the northwestern shore of Galveston Bay. Bechtel is president of the board for the Gulf Coast Protection District, which the state created to collect taxes to help pay for the project. Bechtel questions why any barrier on the islands would be built lower than the nearby Galveston city seawall and the gates spanning the water channel.

"Why are we spending billions of dollars to do that and have it overtopped?" he says. "Let's look at it with common sense."

THERE IS AN ARGUMENT for spending many billions more than the Corps has proposed on protecting Galveston Bay. It rests on a worst-case scenario that envisions hurricane-driven floodwaters battering Houston, the nation's fourth largest city, some 80 kilometers away. Modeling by both Ebersole and Clint Dawson, who studies coastal ocean dynamics at the University of Texas, Austin, shows a Category 4 hurricane could drive a storm surge topping 7 meters into the northern tendrils of the bay, where the shore is lined with sprawling oil refineries and chemical plants.

That's Jim Blackburn's nightmare. Earlier this year, the veteran environmental attorney, co-director of Rice University's Severe Storm Prediction, Education, & Evacuation from Disasters Center (SSPEED), pictured that scenario as he stood on a bluff looking across the placid waters of the bay near Buffalo Bayou. At his back, a concrete tower topped by a single enormous star marked the site of the 1836 Battle of San Jacinto, which paved the way for Texas's independence from Mexico. Ahead, Blackburn faced an industrial landscape of metal towers and white tanks that sprouted like giant mushrooms along the waterfront.

"Every one of these tanks could be hit. That's what I'm horrified about," he said. "This complex is incredibly vulnerable."

About 30% of the country's crude oil and 25% of its natural gas are processed in coastal Texas, much of it around Galveston Bay. The region's refineries can churn out 2.5 million barrels a day, nearly 15% of the U.S. capacity. Oil is stored in more than 4600 huge tanks, many near Houston.

If a 500-year storm hit, Rice scientists have found, the surge and winds could damage some 700 tanks, spilling as much as 470 million liters of oil—almost as much as the 2010 *Deepwater Horizon* spill in the Gulf of Mexico, the largest in U.S. history. The waters could also inundate Houston neighborhoods kilometers upstream of the tanks. "The toxic goo that might be produced could set back not only the region, but the country, for decades," says Houston City Council Member David Robinson, an architect who chairs the council's infrastructure committee.

For Robinson, some version of the Ike Dike is the cornerstone of preventing that nightmare. But he worries it's not enough. That's because, even if the defenses succeed in reducing a storm surge originating from the gulf, a major hurricane could still generate a surge inside the bay.

Houston, surrounding Harris County, the Port of Houston, and a local entrepreneur have committed $1 million to SSPEED for studies. Blackburn and his colleagues have already touted one solution. They propose building a string of artificial islands in the bay—in effect a second wall guarding its northwestern reaches, including Houston. "The Dutch talk about multiple lines of defense," Blackburn says. "That's what we're really talking about here."

One obstacle is cost: Blackburn estimates building the islands would add $3 billion to $6 billion to the price of the plan advanced by the Corps. Another issue is the potential environmental damage caused by yet another huge engineering scheme. Regulators "are just not going to let some massive island be built in the middle of Galveston Bay," says Bob Stokes, president of the Galveston Bay Foundation, an environmental group.

Environmental concerns have already prompted the Corps to ditch one part of the Ike Dike—a proposed gate system across the shallow San Luis Pass, a smaller channel joining the gulf to the bay's western end. Putting a barrier there would "cause irreparable environmental impacts" by interfering with ecologically rich tidelands, Burks-Copes says. Stokes fears that even the remaining gates on the deeper Bolivar Roads channel could meddle with tidal currents and wildlife—including commercially valuable shrimp. As a result, his group has withheld judgment on the agency's plan. "It's hard to say we're for or against it," Stokes says, "until we actually have a deeper level of environmental analysis."

All these protection plans come an important caveat: They would do little to prevent the destruction caused by storms like 2017's Hurricane Harvey, which brought record inland rains—not a wind-driven storm surge—that flooded Houston neighborhoods.

OTHER CITIES in the United States and elsewhere are wrestling with similar issues as they consider new defenses against storm surges. In September, the Corps issued plans for a $52 billion network of barriers in New York Harbor to fend off a repeat of 2012's Hurricane Sandy. That option beat an earlier proposal for a single gated wall guarding the whole harbor, which earned widespread criticism in part for its $119 billion price tag. Also in September, the agency said it would reconsider its proposal to build 10 kilometers of seawalls and gates, 6 meters high in some places, around downtown Miami. Local officials and environmental groups had strongly objected to the plan, fearing it could further degrade Biscayne Bay and ruin the city's famous shoreline.

In Singapore, officials announced plans earlier this year to study the feasibility of storm barriers. Indonesia is considering a giant sea wall to protect Jakarta, its sinking capital, even as it builds a new capital city elsewhere. In 2020, Venice, Italy, after decades of debate and delay, finally inaugurated a system of 72 mobile walls to seal off its lagoon and protect the flood-prone city from rising seas and extreme high tides. The system is working—but environmentalists say it threatens to destroy salt marshes.



Hurricane Harvey's intense rains flooded this chemical facility in 2017. A wind-driven storm surge could be more destructive.

Underlying many of the debates about the costs and impacts of flood barriers are deeper questions: How much risk is tolerable, and what is worth preserving? Do you build for a once-in-a-lifetime or a once-in-a-millennium storm? And how do you account for the uncertain effect climate change will have on those odds?

In the Netherlands, key coastal defenses are built to withstand freakishly rare 10,000-year storms. That partly reflects the existential threat of flooding to a nation in which one-third of the land is below sea level, including much of Rotterdam. In the United States, by contrast, the Corps frequently designs projects for a storm expected to hit once in a century.

There's no reason different regions should adopt the same standard, says engineer Gregory Baecher of the University of Maryland, College Park, who served on a National Academy of Sciences panel in the 2010s that studied how the Corps evaluates risks. In the Netherlands, he notes, surges from a 1000-year storm aren't much deeper than those caused by a 100-year storm, because even very rare storms there aren't much more severe. Gulf Coast hurricanes, however, can dwarf European storms. By one estimate, the surge from a 100-year storm in the gulf is roughly equal to a once-in-a-millennium storm in the Netherlands. So, building to the more conservative Dutch-level standard along the gulf would be an enormous task. "Everybody loves the Dutch," Baecher says. But, "They've got an easier problem."

Merrell says the split also reflects a different attitude. The Dutch build to prevent flooding, whereas the United States is more willing to accept flood damage and then rebuild. But allowing another Ike-scale flood to hit the Texas coast, Merrell says, would be "insane. … We have to protect it. We can't recover it anymore. It's just too expensive."

The Corps says federal law requires it to build projects that, over 50 years, will produce benefits greater than the cost. If a region wants something more robust, it must be willing to pay the higher price—but it might be a lot higher, warns agency engineer Mike Braden. "How close to 100% [protection] are we trying to get?" he asked earlier this year as he stood on the beach next to Burks-Copes. "As we try to approach that 99%, 100% solution, the costs exponentially go up."

Braden, who arrived in Galveston earlier this year, has a title out of a Marvel superhero movie: He's chief of the Mega Project Division at the Corps. That means he's in charge of shepherding the Texas project through the remainder of the planning process and breaking ground. If everything goes smoothly—if Congress appropriates the money, and the project survives challenges from opponents—it will take decades to complete. The Corps now predicts the system could be finished by 2043.

Braden is open to tweaking the design. But he also hears the clock ticking. His biggest fear is that a massive storm rivaling Hurricane Ike or the 1900 disaster will arrive before the work is done. If we don't want to repeat those catastrophes, he said, "We need to go, go, go, go." ■

PHOTO: CHARLIE RIEDEL/AP PHOTO



CALL FOR PAPERS

SPJ
SCIENCE PARTNER JOURNALS

Journal of Remote Sensing

The *Journal of Remote Sensing* is an online-only Open Access Science Partner Journal published in affiliation with **Aerospace Information Research Institute, Chinese Academy of Sciences (AIR-CAS)** and distributed by the **American Association for the Advancement of Science (AAAS)**. Like all partners participating in the Science Partner Journal program, the *Journal of Remote Sensing* is editorially independent from the *Science* family of journals and AIR-CAS is responsible for all content published in the journal. This journal covers multiple research areas that include theory, science, technology of remote sensing, and interdisciplinary research with earth science and information science. Particular topics of interest within the journal include radiative transfer modeling, biogeosciences remote sensing, remote sensing of energy, and more.

Submit your research to the *Journal of Remote Sensing* today!
Learn more at **spj.sciencemag.org/remotesensing**

The Science Partner Journal (SPJ) program was established by the American Association for the Advancement of Science (AAAS), the nonprofit publisher of the *Science* family of journals. The SPJ program features high-quality, online-only, Open Access publications produced in collaboration with international research institutions, foundations, funders and societies. Through these collaborations, AAAS furthers its mission to communicate science broadly and for the benefit of all people by providing top-tier international research organizations with the technology, visibility, and publishing expertise that AAAS is uniquely positioned to offer as the world's largest general science membership society.

Visit us at **spj.sciencemag.org**

@SPJournals **@SPJournals**

OPEN ACCESS

ARTICLE PROCESSING CHARGES WAIVED UNTIL JULY 2023

INSIGHTS



PERSPECTIVES

Various studies show that COVID-19 vaccination temporarily alters menstrual cycles, but the underlying mechanisms require more research.

VIEWPOINT: COVID-19

COVID-19 vaccination and menstruation

COVID-19 vaccination causes small changes to menstruation that quickly resolve

By **Victoria Male**

The rapid development of safe and effective vaccines against COVID-19 has been a triumph of medical science, but vaccines only work if people take them. Although there is extensive evidence that COVID-19 vaccination does not affect fertility, misinformation that it could has been a major source of vaccine hesitancy among young women. As the vaccination program was rolled out to younger age groups, some people noticed menstrual changes after COVID-19 vaccination, and many members of the public found these reports concerning. Research was needed to generate robust data to inform health care professionals and the public about these potential side effects. Menstrual changes have been reported in association with a variety of vaccines, including those against pathogens other than severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), so a secondary aim of this work is to understand the mechanisms by which vaccine-associated menstrual changes could occur.

Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK. Email: v.male@imperial.ac.uk

By April 2022, the Vaccine Adverse Event Reporting System (VAERS) in the United States had received more than 11,000 reports of menstrual changes and unexpected vaginal bleeding after COVID-19 vaccination (*1*). Yellow Card, the equivalent scheme in the United Kingdom, had received more than 50,000 reports (*1*). These schemes are effective at detecting patterns of serious, but rare, adverse events associated with vaccination: Yellow Card detected the rare clotting disorder vaccine-induced immune thrombotic thrombocytopenia (VITT) that is associated with adenovirus-vectored vaccines, and VAERS identified myocarditis as a rare adverse event associated with mRNA vaccines. However, these systems are not designed to detect increased rates of nonserious events that occur commonly. Because menstrual cycles vary naturally, a particular challenge was determining the extent (if any) to which the changes reported could be attributed to COVID-19 vaccination rather than background variation.

Recent studies illustrate the extent to which menstrual cycles vary in the absence of vaccination. The Norwegian Institute of Public Health (NIPH) studied a cohort of 5688 females aged 18 to 30 years who had been recruited to examine other side effects of COVID-19 vaccination. Participants were asked to recall their prevaccination and postvaccination menstrual cycles: 37.8% described at least one aspect as different from their usual experience, even in prevaccination cycles (*2*). Further, a US study that examined changes to cycle length using data from a menstrual cycle tracking app reported a within-individual standard deviation in cycle

PHOTO: LUKAS BARTH/REUTERS PICTURES

length of 4.2 days in unvaccinated individuals (*1*). This underlines the need for formal approaches, including unvaccinated comparison groups, to identify vaccine-associated cycle changes.

The findings of these formal approaches have been remarkably consistent: COVID-19 vaccination is associated with a small increase in menstrual cycle length, but this rapidly resolves. In a study of data from 3959 US-resident users of the menstrual cycle tracking app Natural Cycles, of whom 2403 were vaccinated and 1556 were unvaccinated, there was no effect of the first vaccine dose, the second dose was associated with an increase in cycle length of 0.45 days, and administration of both doses of vaccine in the same cycle was associated with an increase of 2.32 days (*3*). A follow-up study using the same approach in a global cohort of 19,622 people, of whom 14,936 were vaccinated, reported similar results (*4*). Notably, in both studies, cycle length returned to normal within two cycles. Consistently, data from 9652 US residents who tracked their cycles as part of the Apple Women's Health Study identified an increase in cycle length of 0.5 days after the first vaccine dose and 0.39 days after the second dose, with cycle length returning to normal the next cycle (*1*). In the UK, a prospectively recruited cohort of 79 people who recorded their cycles in real time also found a small, but significant, increase in cycle length only in cycles in which a vaccine dose was given (*5*). Among 3858 North American nurses taking part in the Nurses' Health Study 3, which collects data from its participants biannually, vaccination was associated with increased odds (1.48) of reporting a longer menstrual cycle in the next follow-up questionnaire, but this resolved in subsequent follow-ups (*6*).

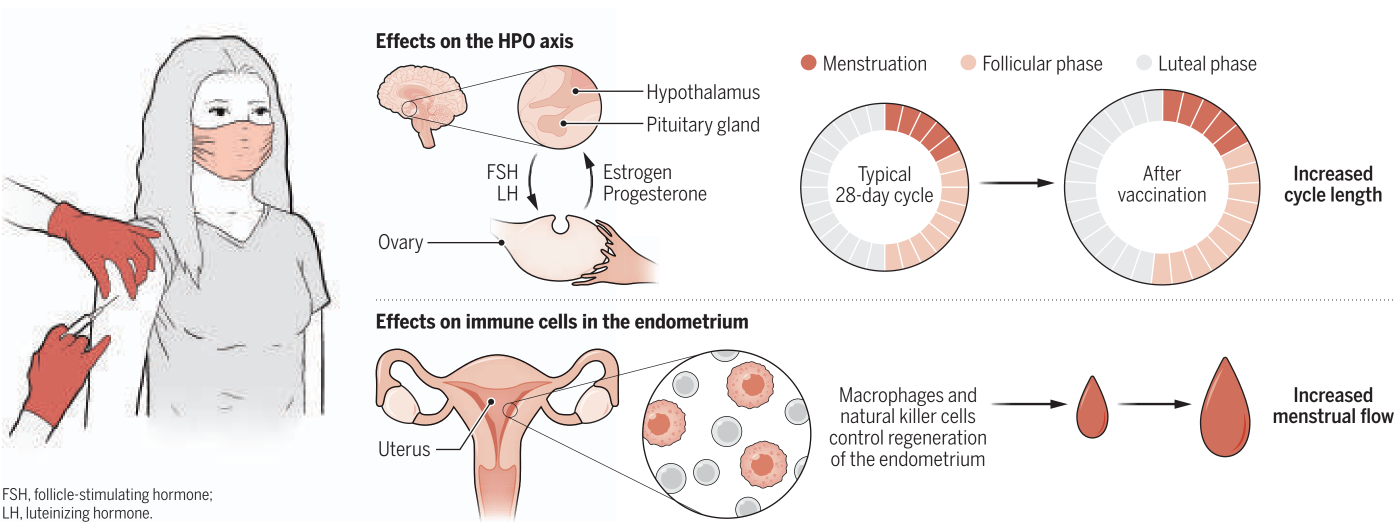
The NIPH study found that heavier-than-normal bleeding was most commonly associated with vaccination, with 13.6% of participants reporting this for the period after they were vaccinated compared with 7.6% for the period before vaccination (*2*). The UK study of 79 participants was unable to detect an increase in menstrual flow associated with vaccination (*5*), although this could be a result of the smaller cohort size. These conflicting results could also reflect the increased risk of recall bias in the NIPH study. Approaches that combine large cohorts and data collection in real time are needed to resolve this.

Given this evidence that COVID-19 vaccination does alter menstrual cycles, albeit temporarily, the obvious question is how? The type of vaccine does not affect the chance that an individual will experience a change to menstrual timing (*1*, *5*–*7*) or flow (*5*, *7*, *8*), suggesting that the effect is a result of the immune response to vaccination rather than a specific vaccine component. In support of this, menstrual changes have previously been reported with typhoid (*9*), hepatitis B (*10*), and human papillomavirus (HPV) (*11*) vaccines. Indeed, as early as 1549, the doctor Wan Chhüan observed that inoculation against smallpox could bring on menstruation unexpectedly (*12*). Two studies have addressed the hypothesis that menstrual changes after COVID-19 vaccination are associated with activation of the immune response, but with conflicting results. The 79 participants in the prospectively recruited UK cohort recorded their experience of common immune-mediated vaccine side effects, and no association was found between the extent of side effects and changes to cycle length or flow (*5*). Conversely, a survey of 27,143 menstruating individuals found that those who experienced fever or fatigue postvaccination were more likely to experience a heavier-than-usual period (*8*). Each of these approaches has weaknesses: The UK cohort is potentially underpowered to detect an association, whereas the survey relies on participants accurately recalling their experiences and is potentially influenced by respondents being more likely to recall other side effects if they noticed a menstrual change.

Two biologically plausible mechanisms by which immune stimulation might cause menstrual changes have been proposed: Innate immune responses could transiently interfere with the hormones that drive the menstrual cycle, or they could affect macrophages and natural killer cells in the lining of the uterus, which control the breakdown and regeneration of this tissue through the cycle (see the figure). In support of the hypothesis that the effects are hormonally mediated, individuals in whom the ovarian hormones estrogen and progesterone are supplied exogenously by combined hormonal contraception are less likely to experience menstrual changes after vaccination (*5*, *7*). Furthermore, the timing of vaccination within the menstrual cycle affects whether cycle length increases. The menstrual cycle is divided into two phases: the follicular phase, which occurs before ovulation and can be prolonged by hormonal alterations, and the luteal phase, which occurs after ovulation and is more consistent in length. If menstrual changes are mediated by immune effects on the control of ovarian hormones, vaccination would be expected to prolong

Possible mechanisms by which vaccines affect menstrual cycles

Cytokines produced in response to COVID-19 vaccination could affect the hormonal dialogue between the hypothalamus, pituitary gland, and ovaries (HPO), lengthening the cycle. Cytokines could also affect macrophages and natural killer cells in the lining of the uterus, which mediate tissue repair, resulting in heavier menstrual flow.



GRAPHIC: A. MASTIN/*SCIENCE*

the follicular phase, but this can only occur if vaccines are administered during this phase. Indeed, the Apple Women's Health Study found that cycle-length increases are only associated with vaccination in the follicular phase of the cycle (*2*).

In support of the possibility that COVID-19 vaccination affects immune cells in the uterine lining, the survey of 27,143 menstruating individuals found that increasing age was associated with an increased risk of heavier bleeding (*8*). This could suggest that altered tissue repair, which is mediated by immune cells in the uterus and may be less effective in older people, is the mechanism by which COVID-19 vaccination increases menstrual flow.

The evidence for the underlying mechanism is therefore mixed and could be consistent with effects mediated by both ovarian hormones (affecting cycle length) and endometrial repair (affecting menstrual flow). Further research is required to definitively identify the pathways involved, but now that there is evidence that COVID-19 vaccination is associated with menstrual changes, more-involved studies that track blood hormone levels before and after vaccination and studies of immune cells isolated from endometrial biopsies or menstrual fluid are justified.

Does SARS-CoV-2 infection affect menstruation? The nature of COVID-19 vaccination makes it amenable to studies that track menstrual cycle parameters before and after exposure, but this is harder for infection because it is unpredictable, it may last for days or weeks, and many people may be unaware that they have been infected, making it more difficult to define an uninfected control group. In studies early in the pandemic, between 15 and 25% of individuals reported changes to their periods after SARS-CoV-2 infection, although one study was on individuals who were hospitalized with COVID-19, one was on individuals with Long Covid, and none included an uninfected control group, so these are likely to be overestimates (*13–15*). More recently, the Nurses' Health Study 3, which compared menstrual cycle length and regularity self-reported in 2011 to 2016 to that self-reported in 2021 found no effect of SARS-CoV-2 infection, although the timing of the questionnaires (not immediately pre- and post-exposure) and the relatively coarse detail in which participants could respond limit the ability of this approach to detect small or temporary effects (*6*).

There are important lessons to be learned. More than half of the world's population menstruate at some time in their lives, yet data about effects on menstruation are rarely collected in vaccine trials. This must change—not least to offer reassurance that this area of public health is also taken seriously by vaccine developers. Furthermore, some information can only be collected in randomized controlled trials. Approaches using menstrual cycle tracking apps have proved powerful because large volumes of data are available and data collection in real time mitigates recall and recruitment bias. However, app users are not representative of the global population because they live mainly in high-income countries and young, white, educated individuals are overrepresented (*1*, *3*, *4*). App users are also aware of being vaccinated, and this may affect their perceptions of aspects of menstruation that are partially or wholly subjective, such as menstrual flow, pain, and premenstrual syndrome (PMS) symptoms. The inclusion of a blinded control group, as in vaccine trials, precludes these problems.

> **"...data about effects on menstruation are rarely collected in vaccine trials."**

But there are also opportunities. The recent advances in defining vaccine effects on the menstrual cycle open several avenues of inquiry. Vaccination is planned and occurs at a single point in time, and thus, importantly, individuals who are already planning to receive a vaccine dose can be recruited, which circumvents the ethical challenges of giving participants an immune stimulus purely for experimental reasons. These individuals can participate in studies not only to fully define how immune stimulation affects female reproductive parameters but also to address the converse question: How do menstrual cycle phase and the use of hormonal contraception affect the immune response? There is an opportunity, finally, to start making real progress in an area that has historically been understudied. ■

REFERENCES AND NOTES

1. E. A. Gibson *et al*., medRxiv 2022.07.07.22277371 (2022).
2. L. Trogstad, SSRN ssrn.3998180 (2022).
3. A. Edelman *et al*., *Obstet. Gynecol*. **139**, 481 (2022).
4. A. Edelman *et al*., *BMJ Med*. **1**, e000297 (2022).
5. A. Alvergne *et al*., *Front. Reprod. Health* **4**, 952976 (2022).
6. S. Wang *et al*., *Am. J. Obstet. Gynecol*. **227**, 739.e1 (2022).
7. A. Alvergne *et al*., medRxiv 2021.11.23.21266709 (2022).
8. K. M. N. Lee *et al*., *Sci. Adv*. **8**, eabm7201 (2022).
9. A. R. Lamb, *Arch. Intern. Med*. **XII**, 565 (1913).
10. T. Shingu *et al*., *Kurume Med. J*. **29**, 123 (1982).
11. S. Suzuki, A. Hosono, *Papillomavirus Res*. **5**, 96 (2018).
12. J. Needham, *East. Horiz*. **19**, 6 (1980).
13. K. Li *et al*., *Reprod. Biomed. Online* **42**, 260 (2021).
14. S. M. Khan *et al*., *Am. J. Obstet. Gynecol*. **226**, 270 (2022).
15. H. E. Davis *et al*., *EClinicalMedicine* **38**, 101019 (2021).

ACKNOWLEDGMENTS

The author receives financial support from Borne, a charity that funds preterm-birth research.

10.1126/science.ade1051

GEOLOGY

Defining the onset of the Anthropocene

Twelve sites are considered for defining the Anthropocene geological epoch

By **Colin N. Waters**[1] *and* **Simon D. Turner**[2]

Earth's geological history is divided into chronostratigraphic units that distinguish phases in the planet's evolution by summarizing complex biotic, geochemical, and climatic changes. Over the past century, many components of the Earth system have changed so much that they no longer occur within the ranges evident during the Holocene—the geological epoch that represents the past ~11,700 years. There are also distinct geological traces that warrant recognition as a new geologic epoch: the Anthropocene. The Anthropocene Working Group (AWG), a task group of the Subcommission on Quaternary Stratigraphy (SQS) of the International Commission on Stratigraphy (ICS), have been working to decide precisely when the Anthropocene began, with a focus around the mid-20th century. The definition will need to identify specific physical properties in sediment layers, or strata, that capture the effects of recent increases in human population; unprecedented industrialization and globalization; and changes imposed on the landscape, climate, and biosphere (*1–7*).

The definitions of chronostratigraphic units form the basis of the International Chronostratigraphic Chart (ICC) (*8*) and standardize the geological time scale—e.g., when specific periods, epochs, and ages begin and end and how they can be identified in strata. Protocols established by the ICS to formalize chronostratigraphic units require the definition of a global boundary stratotype section and point (GSSP). This requires the selection of a single reference point for defining the base (the lowermost part) of a chronostratigraphic unit, from which correlation of an isochronous boundary (i.e.,

[1]School of Geography, Geology and the Environment, University of Leicester, Leicester, UK. [2]Department of Geography, University College London, London, UK. Email: cw398@leicester.ac.uk

one that is the same age everywhere) for the Anthropocene can be achieved globally (*9*). Given the diversity of geological modification that has occurred in different environments, it is challenging to find a single GSSP that can best represent a global reference for the Anthropocene.

A candidate Anthropocene GSSP site must have at least one primary stratigraphic marker to be used as a reference to correlate the boundary at the selected GSSP site to other locations across the planet (see the figure). These markers include those that indicate a geochemical change—e.g., plutonium isotopes, carbon-14 (^{14}C), and $\delta^{15}N$ [based on the ratio of nitrogen-14 (^{14}N)/^{15}N isotopes]—or the appearance of anthropogenic particles—e.g., spheroidal carbonaceous particles (SCPs), which are a type of fly ash only produced by high-temperature combustion of coal or fuel oil (see the photo). Ideally, the site should have precisely datable strata with enough resolution to fix the onset of the Anthropocene to a specific year (*10*). The strata, mainly extracted as cores at the sites, must represent a complete time record, with no gaps or disturbances in the critical Holocene-Anthropocene boundary interval. A continuous part of the entire core will need to be conserved (for future reference), and the site should preferably be sufficiently accessible that additional cores can be extracted if needed.

Building on many years of previous research (*10*), 12 candidate sites have been proposed for consideration by the AWG. In addition to the globally important stratigraphic markers (plutonium isotopes, ^{14}C, $\delta^{15}N$, and SCPs), numerous regional markers have been analyzed, including microplastics, heavy metals, and even biological proxies such as the population trends of introduced species (neobiota). AWG voting members are tasked with selecting one candidate GSSP site.

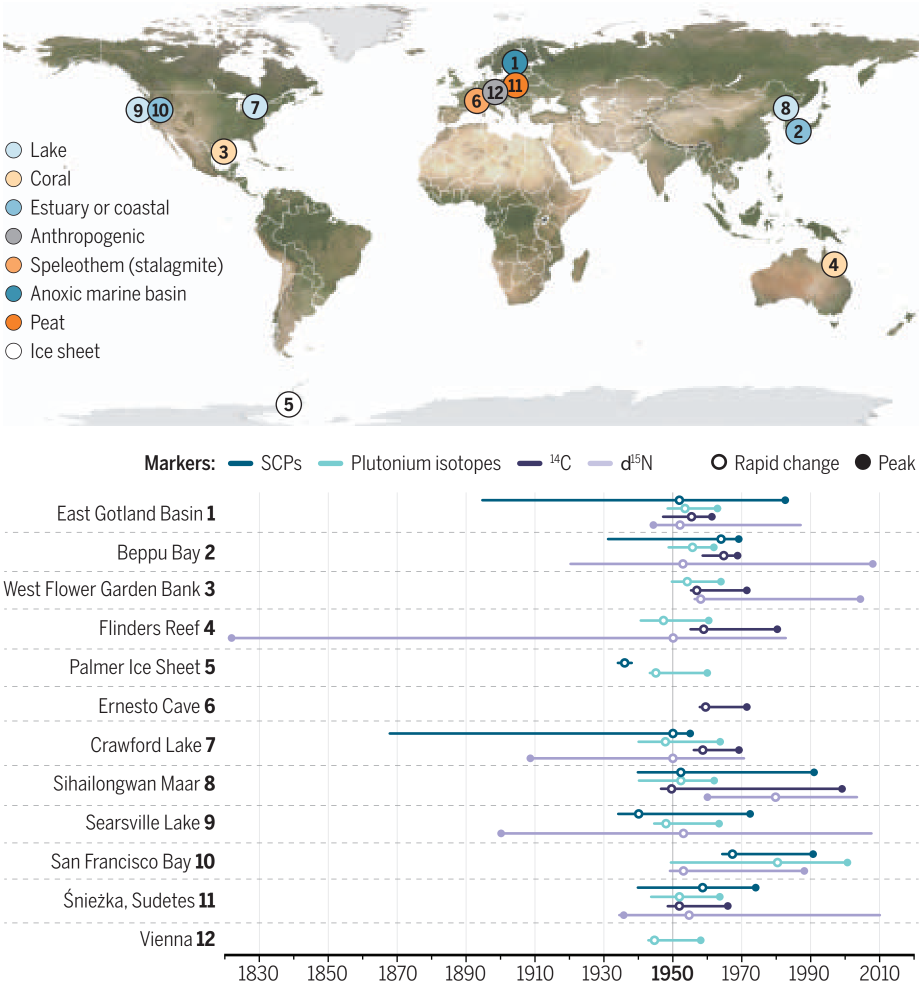
Two of the candidate sites are in marine sediments—one in the Baltic Sea and one in the coastal Beppu Bay of Japan. For the Anthropocene, suitable marine sediments should exhibit sufficient accumulation rates, be in contact with anoxic bottom waters to limit bioturbation (disturbance by living organisms), and be unaffected by "anthroturbation"—i.e., disturbance by human activities, such as trawler fishing (*10*). Both the Baltic Sea and Beppu Bay sites contain dark, carbon-rich silts and clays tens of centimeters thick, with the Beppu Bay site also showing seasonal layering. At both sites, these strata lie above the paler, weakly laminated Holocene sediments, and human activity–driven eutrophication (i.e., nutrient enrichment causing phytoplanktonic blooms) is reflected in the prominent textural and compositional changes in the strata along with several other Anthropocene markers, including $\delta^{13}C$ (based on the ratio of $^{13}C/^{12}C$), $\delta^{15}N$, SCPs, microplastics, and pesticides.

Two candidate sites are from coral reefs—the West Flower Garden Bank (Gulf of Mexico) and the Flinders Reef (Coral Sea, Australia). Here, the skeletons of long-lived corals preserve visible seasonal growth bands that contain geochemical markers of environmental change, which can be resolved precisely back to the 1750s and 1700s, respectively, allowing for suitable comparison of Anthropocene and late Holocene successions. Both sites show clear effects of increased global fossil fuel combustion driving $\delta^{13}C$ to more negative values and proxies indicating warming sea surface temperatures.

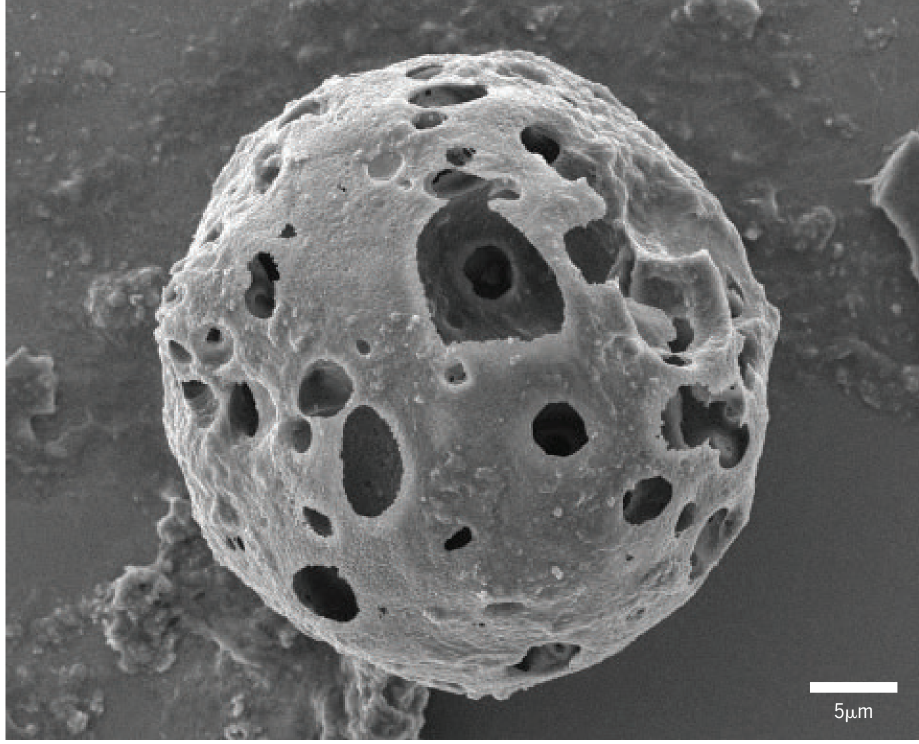
Three lake sites have been proposed—Crawford Lake in Canada, which is meromictic (i.e., its water layers do not intermix); Sihailongwan Lake in China, which is a crater lake with a seasonally stratified water column with a lowermost anoxic layer; and Searsville Lake in the US, created by a dam built in 1892. The Crawford Lake core indicates ecological effects from indigenous occupation followed by European colonization and shows major ecological changes in its microfossils from the 1950s onward, with carbon and nitrogen cycle perturbations and SCP influx. Sihailongwan Lake is more remote and lacks obvious anthropogenic signals during the Holocene but shows the increased concentrations of soot (particulates formed as a high-temperature condensate), SCPs, heavy metals, and polycyclic aromatic hydrocarbons resulting from fossil fuel burning and industrial emissions. Searsville Lake has silts impounded through seasonal storm events with a layer-count chronology tied to local pollution and two earthquake events. At Searsville Lake, SCPs, polychlorinated biphe-

Candidate sites for defining the Anthropocene

Core samples from the 12 candidate sites contain various markers of geochemical change and anthropogenic influence. Globally important stratigraphic markers from each core sample include spheroidal carbonaceous particles (SCPs), plutonium isotopes, carbon-14 (^{14}C), and $\delta^{15}N$ [based on the ratio of nitrogen-14 (^{14}N)/^{15}N isotopes]. Lines show the full range of each marker's presence in the core. The point of rapid change of multiple markers provides chronological correlation between the sites.



GRAPHIC: K. FRANKLIN/*SCIENCE*



A scanning electron microscopy image shows an ~40-μm-diameter spheroidal carbonaceous particle (SCP) from Crawford Lake, Canada.

PHOTO: SARAH ROBERTS

nyls, and pesticides mark the Anthropocene.

The Palmer ice core from the Antarctic Peninsula has also been proposed. Ice cores have previously been selected to define a GSSP for two Holocene subdivisions (*11*). The length of the core from the Palmer site that corresponds to the Anthropocene is by far the longest among the candidates, at ~32 m. This provides more material to analyze per year, which is important given the low concentrations of some markers in this very remote setting. The core displays seasonally resolvable laminae and records snowfall extremes during the Anthropocene, acceleration of atmospheric methane concentrations, and even small amounts of SCPs.

A core from a raised ombrotrophic peat bog—i.e., one that receives nutrients and water from the atmosphere for growth—near the Śnieżka peak in the Sudetes Mountains, Poland, has been proposed. The peat from the site shows an increase of fly ash from thermal power stations, coinciding with the appearance of pollen from introduced plants and perturbations in $\delta^{13}C$ and $\delta^{15}N$, respectively, that align with the 1950s.

Three proposals have proved less favorable as the GSSP site: the Ernesto Cave in Italy, the San Francisco Bay estuary, and a section in anthropogenic deposits in central Vienna. The Ernesto Cave site, based on a core of a stalagmite only 4 mm thick that accumulated during the Anthropocene, has ^{14}C and sulfur records that lag behind monitored atmospheric records by 1 to 2 decades as a result of vegetation and soil delaying atmospheric signals reaching the cave. The site was discounted because of this delay. The San Francisco Bay estuary was originally proposed because it has one of the best-resolved historical records of neobiotic species, more than half of which were introduced after 1960. The core records increases in mercury and SCPs in the 1960s and a paleontological succession of five neobiotic species in the 1980s, with their stratigraphic appearances mirroring their known arrival dates. However, because the age of the succession is relatively poorly constrained and the core stopped before reaching the key interval at the base of the Anthropocene, this site was also discounted as a GSSP. Finally, a section of anthropogenic deposits in central Vienna reflects the city's development and can be dated by artifacts, including debris from World War II. It provides a near-200-year history of pronounced urban changes. However, because of notable gaps during this time interval, the Vienna site is also not suitable as a GSSP.

Together, the 12 sites provide a diverse perspective on the geological reality of the Anthropocene. Although showing regional signals from industrial and agricultural processes with decade-scale variations in timing, global signals provide clear instances of time that can be linked between the sites. Hence, the base of strata marking the beginning of the Anthropocene can be recognized and correlated between all of the sites, whether on land, in ice, or beneath the sea. For instance, one such marker is the acute influx of plutonium isotopes created by a series of aboveground hydrogen bomb test detonations commencing in late 1952 (*12*). The plutonium provides a sharper global marker than those produced by larger and longer-lasting drivers of Earth system change—e.g., increased burning of fossil fuels and expansion of agriculture—and consequently has been selected by many of the sites as the primary marker.

The AWG will select their preferred candidate GSSP site based on comparison of the robustness of the core chronology and clarity of physical markers around the onset of the Anthropocene. The chosen site will be used to compose the age name. For example, if the Crawford Lake is chosen, then the current age may be named the Crawfordian Age, which would be the first age of the Anthropocene epoch. The chosen GSSP level at the site will be used to set the precise beginning of the Anthropocene.

The AWG will vote on the site in late 2022, which is only the initial stage in a process. It will be followed by a proposal to be voted on by the SQS, then voted on by the ICS, and finally ratified by the International Union of Geological Sciences (IUGS). Only then will the Anthropocene be formally approved as a component of the ICC. Each of these votes is required to pass by a 60% majority and may fail at any stage, in which case it cannot be subsequently modified for at least 10 years. If approved, the Anthropocene will have a fixed scientific definition with a precise start date that is characterized by a section in the strata with well-defined markers. This will help focus studies on the scale and effect of recent human activity on the planet, contrasting with the relative stability of the preceding Holocene. ■

REFERENCES AND NOTES

1. W. Steffen, W. Broadgate, L. Deutsch, O. Gaffney, C. Ludwig, *Anthr. Rev*. **2**, 81 (2015).
2. M. J. Head *et al*., *Episodes* 10.18814/epiiugs/2021/021031 (2021).
3. C. N. Waters *et al*., *Science* **351**, eaad2622 (2016).
4. J. Zalasiewicz, C. N. Waters, M. Williams, C. P. Summerhayes, Eds., *The Anthropocene as a Geological Time Unit: A Guide to the Scientific Evidence and Current Debate* (Cambridge Univ. Press, 2019).
5. J. Syvitski *et al*., *Commun. Earth Environ*. **1**, 32 (2020).
6. J. Zalasiewicz, C. Waters, M. Williams, in *A Geologic Time Scale 2020*, F. M. Gradstein, J. G. Ogg, M. D. Schmitz, G. M. Ogg, Eds. (Elsevier, 2020), pp. 1257–1280.
7. M. Williams *et al*., *Palaeontology* **65**, e12618 (2022).
8. K. M. Cohen, S. C. Finney, P. L. Gibbard, J.-X. Fan, The ICS International Chronostratigraphic Chart (IUGS, 2022); https://stratigraphy.org/ICSchart/ChronostratChart2022-02.pdf.
9. A. Salvador, *International Stratigraphic Guide: A Guide to Stratigraphic Classification, Terminology, and Procedure* (IUGS and Geological Society of America, 1994).
10. C. N. Waters *et al*., *Earth Sci. Rev*. **178**, 379 (2018).
11. M. Walker *et al*., *Episodes* **41**, 213 (2018).
12. C. N. Waters *et al*., *Bull. At. Sci*. **71**, 46 (2015).

ACKNOWLEDGMENTS

C.N.W. and S.D.T. are, respectively, the Chair and Secretary of the AWG. The authors thank J. Zalasiewicz and M. Head for their comments; the Haus der Kulturen der Welt, Berlin, for its support and collaboration with the AWG and for the many teams involved in compiling the results for the 12 sites; and S. Roberts for supplying the SCP image.

10.1126/science.ade2310

GENETICS

Potential bias in genetic correlations

Mating patterns across two traits can inflate estimates of genetic overlap

By **Andrew D. Grotzinger**[1,2] *and* **Matthew C. Keller**[1,2]

Genome-wide association studies (GWASs) identify genetic variants associated with a trait. Most traits are associated with thousands of variants, and many variants are pleiotropic, meaning they are associated with multiple traits. Pervasive pleiotropy makes it impractical to assess genetic overlap between two traits by tallying the shared variants. For example, traits such as major depression and anxiety are likely associated with a shared set of thousands of variants. Genetic correlation (r_g) estimated from GWASs of a pair of traits is typically interpreted as an overall measure of genetic overlap, providing a useful metric for quantifying shared biology between traits. On page 754 of this issue, Border *et al.* (*1*) report simulation-based and empirical findings that challenge this interpretation.

Border *et al.* investigate the effect on r_g of cross-trait assortative mating (xAM), which occurs when individuals scoring highly for trait Y mate with partners who score high (or low) for a separate trait, Z. The level of xAM will vary depending on the pair of traits being considered and is quantified in their study using cross-trait correlations across spouses. They show that xAM can increase r_g in two ways (see the figure). They demonstrate that xAM will increase r_g due to genetic variants for trait Y being coinherited with the variants for trait Z (and vice versa). This first potential effect of xAM on r_g estimates has been understood for decades as an interpretive caveat in family-based studies (*2*–*4*), although it is perhaps not widely discussed. The authors point out that this first effect of xAM does not cause an actual bias in r_g estimates because the polygenic scores—the cumulative set of genetic variants that affect each trait—are legitimately correlated. This is despite there being no pleiotropic genetic variants that affect both traits. If this were the only issue, the interpretation of r_g as strictly an index of shared biology would require updating, but r_g would still be a valuable metric with etiological and clinical implications for understanding risk for trait Y given trait Z.

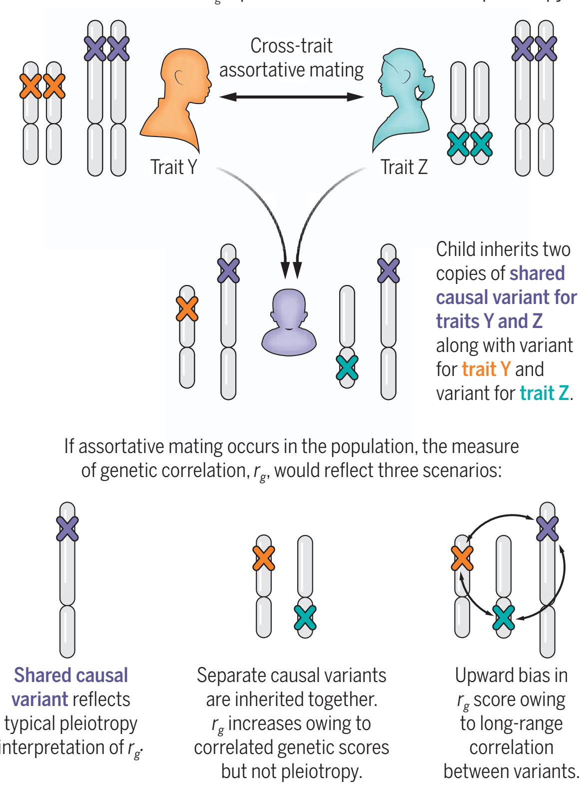
The second effect that xAM has on r_g is potentially more problematic. Commonly used methods for estimating r_g, such as linkage disequilibrium (LD) score regression (*5*), Haseman-Elston regression (*6*), and genomic residual maximum likelihood (REML) (*7*), assume that there is no long-range correlation across causal variants. Border *et al.* find that xAM violates this assumption by inducing long-range correlations between causal variants that are on average positive or negative (i.e., sign-consistent). This causes an upward bias in r_g estimates after even a single generation of xAM. Unlike in the first case where only the interpretation of r_g would shift, this second issue causes bias in the true sense of the term: The inflation of r_g is a statistical artifact that distorts the correlation between polygenic scores. Somewhat reassuringly, the first effect of xAM on r_g will typically be greater than the influence of the second effect, particularly when genomic REML is used.

In the worst-case scenario, xAM for two traits with no overlapping causal variants will produce r_g estimates that are not due to pleiotropy and are upwardly biased. Border *et al.* use simulation findings and empirical r_g and cross-mate, cross-trait correlation estimates to show that, for example, r_g estimates between alcohol use disorder–schizophrenia or height–educational attainment may fall into this worst-case scenario. However, for other pairs of traits, the r_g estimates are too high relative to the estimated levels of xAM to realistically reflect a by-product of xAM alone. Within psychiatric phenotypes, these trait pairs with large r_g estimates include bipolar disorder–schizophrenia and major depressive disorder–anxiety. In addition, estimates of r_g between traits that show little or no xAM, such as certain diseases or late-onset disorders, are still likely to reflect pleiotropy. Even for traits that do show evidence for xAM, r_g will typically reflect some combination of three alternative contributions—pleiotropy, correlated polygenic scores not due to pleiotropy, and bias due to long-range correlation. Therefore, the findings of Border *et al.* should not be interpreted as implying that r_g estimates are wholly unrelated to pleiotropy.

Their results for case-control traits, which compare individuals with a disorder (a case) to those who do not have the disorder (a control), should also be interpreted with caution. As is standard, Border *et al.* quantify xAM between case-control traits using tetrachoric correlations, which estimate the correlation between the continuous risk scores assumed to underlie two disorders. Because patterns of xAM across all levels of genetic risk affect r_g, this approach assumes that the same degree of xAM inferred from mating patterns for cases with extreme risk holds across all levels of risk,

Measuring genetic correlations

Cross-trait assortative mating could result in genetic correlations with certain traits that are explained by three scenarios, two of which inflate the score, r_g, upward even in the absence of pleiotropy.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

[1]Institute for Behavioral Genetics, University of Colorado at Boulder, Boulder, CO, USA. [2]Department of Psychology and Neuroscience, University of Colorado at Boulder, Boulder, CO, USA. Email: andrew.grotzinger@colorado.edu

including for mates who are "subthreshold" for a disorder. However, increased mating between individuals who receive a diagnosis may occur for reasons, such as an elevated chance of meeting a partner in a clinical setting, that have no bearing on mating that occurs between individuals with subthreshold risk. Thus, tetrachoric correlations between disorders, and perhaps especially between psychiatric disorders, may overstate the influence of xAM on r_g estimates.

The findings of Border *et al.* make it clear that more realistic models for why mates correlate within and between multiple traits need to be developed and tested. It may be that xAM occurs only on a few traits of central importance to mating and that most other mate correlations are consequences of the phenotypic and genetic correlations between those central traits and the traits that are actually measured (indirect AM). It is also possible that some xAM estimates are due to mates becoming more similar over time (convergence). Assortment can also reflect social homogamy, which occurs when mate choice is based on environmental components of the trait that are not due to genetics. This could occur, for example, if mates are chosen on the basis of religious beliefs and religion is both due to the environment and affects the traits being studied. Convergence and social homogamy are not expected to affect r_g estimates. Deriving more complete models of the mechanisms through which observed levels of xAM manifest will be important for obtaining a better understanding of downstream effects on r_g.

Border *et al.* demonstrate that xAM is likely to be pervasive and affect many r_g estimates to some degree. This bias will inherently carry forward to results from any methods that use genetic correlations as input, such as Mendelian randomization (*8*) or genomic structural equation modeling (*9*). Therefore, complex trait genetics ignores these problems at its peril. These issues can be addressed by increased care in interpreting r_g as well as through the development of methods that can disentangle the various contributions to r_g estimates. ■

REFERENCES AND NOTES

1. R. Border *et al.*, *Science* **378**, 754 (2022).
2. L. J. Eaves, A. C. Heath, N. G. Martin, *Behav. Genet.* **14**, 371 (1984).
3. M. C. Keller *et al.*, *PLOS Genet.* **9**, e1003451 (2013).
4. G. P. Vogler, J. C. DeFries, *Behav. Genet.* **15**, 111 (1985).
5. B. Bulik-Sullivan *et al.*, *Nat. Genet.* **47**, 1236 (2015).
6. J. K. Haseman, R. C. Elston, *Behav. Genet.* **2**, 3 (1972).
7. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, *Am. J. Hum. Genet.* **88**, 76 (2011).
8. F. P. Hartwig, N. M. Davies, G. Davey Smith, *Genet. Epidemiol.* **42**, 608 (2018).
9. A. D. Grotzinger *et al.*, *Nat. Hum. Behav.* **3**, 513 (2019).

10.1126/science.ade8002

ORGANIC CHEMISTRY

Adding functions to pyridines

Chemical reactions break a pyridine ring to allow its modification

By **Jung Min Joo**

Pyridine (C_5H_5N) is a launch point for creating a wide range of chemicals, including those used in drug discovery, catalysis, and materials science (*1*). It consists of a hexagonal ring of five carbon-hydrogen (C–H) pairs and one nitrogen (N) atom. The synthesis of pyridine derivatives bearing substituents other than hydrogen requires custom-designed processes. Among them, substitutions of one substituent for another are frequently performed while maintaining the integrity of the ring. However, the efficiency of these methods and the diversity of pyridine compounds that they yield need improvement (*2*). On pages 773 and 779 of this issue, Boyle *et al.* (*3*) and Cao *et al.* (*4*), respectively, report different approaches to breaking a pyridine ring, replacing the hydrogens, and then restoring the ring. These techniques present possible new pathways to generate a wider variety of agrochemicals, pharmaceuticals, and material compounds.

> "...new pathways... generate a wider variety of agrochemicals, pharmaceuticals, and material compounds."

The effects of substituents that replace the hydrogens in a pyridine is an important topic in organic chemistry. Halogenation—the replacement of these hydrogens with a halogen atom—is useful as an intermediate step for making pyridine derivatives because the installed halogen can be much more readily substituted by another atom or group of atoms as compared with hydrogen (*5*). In addition, the halogen substituents can facilitate intermolecular interactions that are essential for certain desired functions such as binding to target proteins (*6*).

Halogenation of pyridines can be accomplished through their reaction with electrophilic chemicals (*7*). These electrophiles are electron poor and can form chemical bonds by accepting an electron pair from electron-rich chemicals. However, because the nitrogen atom attracts electrons from carbon atoms of the ring, the carbon atoms themselves are electron poor and thus do not react easily with electrophiles. Furthermore, the nitrogen atom reacts preferentially with electrophilic reagents to form the corresponding pyridinium salts, which are more electron poor than pyridines and thereby less likely to react with electrophilic halogen species. As a result, halogenation requires harsh reaction conditions, usually involving strong acids and elevated temperatures, to generate very strong electrophiles that can enable the reaction. This limits the range of reactants that can be used and the products that can be created.

Positional selectivity is another important consideration in preparing pyridine derivatives. Although the most relatively electron-rich positions on the ring (positions 3 and 5) allow electrophilic substitution, it is difficult to control which of these positions is substituted or whether one or both of the positions are substituted. In addition, some substituents could alter the electronic and structural characteristics of the pyridine ring in ways that enable chemical functions but may decrease the positional selectivity of the reaction on the ring (*7*). There are also certain products that require the pyridine to react while positioned alongside other aromatic rings (such as those with delocalized electrons) that would compete with the aromatic pyridine ring for halogenation.

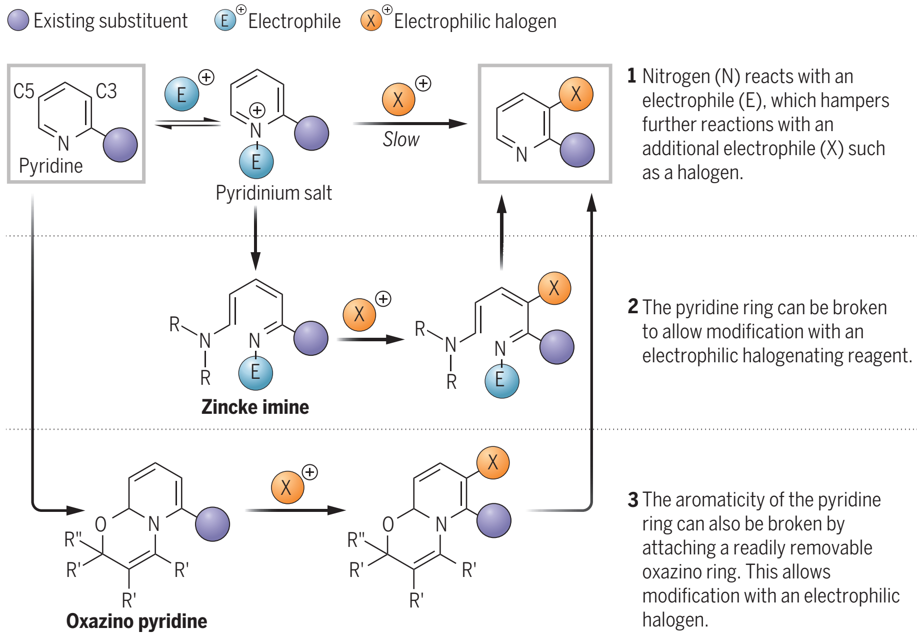
Different strategies addressing the halogenation of pyridines present advantages and disadvantages. For example, the use of bases instead of acids to mediate halogenation tends to generate more reactive intermediates (*8*). Some indirect methods use more reactive and accessible intermediates to avoid the use of strong electrophiles (*9, 10*). However, the requirements of these other strategies can limit the range of possible substituents. Thus, more versatile means are needed to prepare halogenated pyridines from simple pyridines or complex substituted pyridines.

Boyle *et al.* developed a strategy that transforms the electron-deficient pyridine into an electron-rich molecule, enabling it to react with electrophilic halogenat-

Department of Chemistry and Chemistry Institute for Functional Materials, Pusan National University, Busan 46241, South Korea. Email: jmjoo@pusan.ac.kr

A sequence of open and closed rings

Generating halogen derivatives of pyridine is typically a slow and inefficient process. However, two approaches that break open the pyridine ring allow it to readily react with a reagent and with high positional selectivity on the ring. The ring can then be restored. Both pathways allow halogenation, and it may be possible to explore other ring modifications. C3 and C5 are positions on the ring.



GRAPHIC: N. CARY/*SCIENCE* BASED ON J. M. JOO

ing reagents with high site selectivity (see the figure). The authors were inspired by a century-old method that creates reactive intermediates by opening the pyridine ring. The process, called a Zincke reaction, was originally devised to convert pyridines to pyridinium salts by substituting at the nitrogen site (*11*). The intermediate compound during this reaction, known as the Zincke imine intermediate, has one of the carbon-nitrogen (C–N) bonds broken, thus opening the ring. Boyle *et al.* adopted the Zincke reaction for installing substituents at carbon atoms of pyridines by using an electrophilic reagent, trifluoromethanesulfonic anhydride $[(CF_3SO_2)_2O]$, and an amine additive, dibenzylamine $[(C_6H_5CH_2)_2NH)]$, to produce electron-rich Zincke imine intermediates. The resultant intermediates are even more electron rich than conventional electron-rich aromatic rings and can outcompete the latter for electrophilic halogenation. Subsequent ring closure to regenerate the pyridine ring was also unexpectedly easy and did not cause the highly reactive Zincke imine intermediates to decompose through undesirable side reactions. This reaction sequence of ring-opening, halogenation, and ring-closing allowed the authors to synthesize several different C3-halogenated pyridine derivatives.

Cao *et al.* took a different path to produce the same result by transforming pyridines to electron-rich oxazino $(C_4N_1O_1)$ pyridine intermediates. The oxazino pyridine can react with radical species (compounds with at least one unpaired electron) as well as with electrophilic reagents to substitute hydrogen with another atom or group of atoms. Similar to the intermediates described by Boyle *et al.*, the formation of oxazino pyridine intermediates is well studied (*12*), but the potential for their use in pyridine functionalization has not been explored. Cao *et al.* used inexpensive reagents, including dimethyl acetylenedicarboxylate $(CH_3OCOC\ CCO_2CH_3)$ and methyl pyruvate $(CH_3COCO_2CH_3)$, to break the aromaticity of the pyridine ring and make it more reactive. This approach is similar to other known reactions that break the aromaticity of the pyridine ring (*13*). However, these other pathways tend to generate intermediates that are either too stable and do not allow restoration of the pyridine ring after substitution or are too unstable and restore the ring before the desired substitutions can happen. By attaching a readily removable oxazino ring, Boyle *et al.* created stable intermediates compared with these other approaches. They demonstrated the compatibility of their method using a range of reagents, including trifluoromethyl (CF_3), perfluoroalkyl (C_nF_{2n+1}), halogen (chlorine, bromine, and iodine), nitro (NO_2), sulfanyl (SC_6H_5), and selenyl (SeC_6H_5) group donors. After the substitutions, the attached oxazino ring was removed by means of acid treatment to yield the final product.

The approaches developed by Boyle *et al.* and Cao *et al.* can be performed in a "one-pot" process that only requires the sequential addition of reagents without changing the reaction vessel. Both methods exhibit high selectivity for the pyridine ring, which allows the selective halogenation of the ring without halogenating other aromatic rings. They also have high selectivity for the C3 position within the pyridine ring. Although the two electron-rich positions, C3 and C5, exist in the intermediates of both syntheses, C3 is preferentially substituted over C5 except in a few cases, such as when a substituent is already present at C3. Both approaches also allow sequential halogenation, which produces pyridines with two halogen atoms in the ring—first at C3 and then at C5.

The techniques of Boyle *et al.* and Cao *et al.* offer opportunities to explore pyridine functionalization. The preferential halogenation of pyridines over other electron-rich aromatic rings, with high positional selectivity, removes the constraints of having to generate halogen-containing pyridines at an early stage of synthesis to prevent halogenation of all aromatic rings. Furthermore, selective halogenation can be combined with the diversification of carbon-halogen bonds to produce pyridine derivatives with varying substituents. Apart from electrophilic halogenating reagents, other electrophiles that are not typically amenable to pyridine functionalization may now be considered. The ease of producing reactive intermediates with a wide range of installable substituents should optimize the creation of drug candidates and organic functional materials that were not achievable in the past. ■

REFERENCES AND NOTES

1. C. H. McAteer, R. Murugan, J. H. Yamamoto, in *Comprehensive Heterocyclic Chemistry IV*, D. S. Black, J. Cossy, C. V. Stevens, Eds. (Elsevier, 2022), pp. 217.
2. K. Murakami, S. Yamada, T. Kaneda, K. Itami, *Chem. Rev*. **117**, 9302 (2017).
3. B. T. Boyle, J. N. Levy, L. de Lescure, R. S. Paton, A. McNally, *Science* **378**, 773 (2022).
4. H. Cao, Q. Cheng, A. Studer, *Science* **378**, 779 (2022).
5. I. J. S. Fairlamb, *Chem. Soc. Rev*. **36**, 1036 (2007).
6. G. Cavallo *et al*., *Chem. Rev.* **116**, 2478 (2016).
7. M. R. Grimmett, in *Advances in Heterocyclic Chemistry*, A. R. Katritzky, Ed. (Academic Press, 1993), vol. 58, pp. 271.
8. M. Balkenhohl, P. Knochel, *SynOpen* **2**, 78 (2018).
9. S. P. Zucker, F. Wossidlo, M. Weber, D. Lentz, C. C. Tzschucke, *J. Org. Chem*. **82**, 5616 (2017).
10. J. S. Wright, P. J. H. Scott, P. G. Steel, *Angew. Chem. Int. Ed*. **60**, 2796 (2021).
11. T. Zincke, G. Heuser, W. Möller, *Justus Liebigs Ann. Chem*. **333**, 296 (1904).
12. R. Huisgen, M. Morikawa, K. Herbig, E. Brunn, *Chem. Ber*. **100**, 1094 (1967).
13. X.-Y. Zhou, M. Zhang, Z. Liu, J.-H. He, X.-C. Wang, *J. Am. Chem. Soc*. **144**, 14463 (2022).

ACKNOWLEDGMENTS

The National Research Foundation of Korea (NRF-2022R1A2C2008629) is gratefully acknowledged.

10.1126/science.ade5501

POLARITON CHEMISTRY

Using mirrors to control molecular dynamics

An optical cavity mixes molecular vibrations with light and changes chemical reactivity

By **Lev Chuntonov**

Polariton chemistry is an emerging field that explores whether the products and rates of chemical reactions can be controlled by placing a reaction mixture between a pair of mirrors (an optical cavity), allowing light to interact with the molecules through molecular vibrational excitation. In this setup, molecules absorb energy from the light and release it back to the cavity many times over. Despite recent demonstrations (*1*–*3*), the details of this phenomenon are poorly understood (*4*). On page 790 of this issue, Chen *et al*. (*5*) describe the molecular dynamics of iron pentacarbonyl [$Fe(CO)_5$] in an optical cavity. By using ultrafast infrared spectroscopy with subpicosecond time resolution, the authors observed how the cavity affects the intramolecular energy dissipation and rearrangement of CO moieties around the Fe atom. The findings add to our understanding of how polariton chemistry might be used to perform molecular transformations that are intractable by conventional synthesis.

Chemical reactions depend on how a molecule shares energy between different vibrations (imagine a molecule changing the way it vibrates from stretching a chemical bond to bending) (*6*). Such intramolecular vibrational relaxation (IVR) can be important when considering polariton chemistry (*7*). When the vibrations of many molecules interact with light trapped inside an optical cavity, the molecules and the cavity are said to be strongly coupled because the energy excess in one component can be readily distributed to the other. This coupling can be described by a quantum-mechanical state shared by the combined system (the molecules and the cavity) known as a vibrational polariton.

Reaction coordinates describe the pathways that reactants take as they transform into products. Vibrations and the reaction coordinate may interact directly, for example, when the strong vibration of a chemical bond leads to its cleavage (*1*), or indirectly, when the vibrational energy in one bond is transferred through other bonds to the one that is cleaved (*3*). Although such intramolecular vibrational couplings in the reagents, products, and intermediates affect reactions, identifying the corresponding mechanisms in detail is difficult. Learning how coupling to the cavity changes the way vibrational energy is dissipated can help in developing cavity-catalyzed chemistry for other reactions.

Schulich Faculty of Chemistry and Solid-State Institute, Technion–Israel Institute of Technology, Haifa, Israel. Email: chunt@technion.ac.il

Highly symmetrical $Fe(CO)_5$ undergoes an ultrafast process of pseudo-rotation in which the CO moieties slightly change their angles in such a way that the molecule appears to have rotated by 90°, even though it did not rotate at all (*8*). Chen *et al.* measured the directional change of the C–O stretching vibration using infrared spectroscopy (*9*) and found that the strong coupling between the C–O vibrations and the cavity slowed down the pseudo-rotation rate by about one-fourth of the corresponding value for reactions that occur outside the cavity. At the same time, the rate of energy redistribution between different C–O vibrations nearly doubled at room temperature (25°C) and increased fourfold at 80°C. These findings apparently contradict a recent conclusion that the pseudo-rotation reaction coordinate and the C-O stretching vibrations do not interact strongly (*10*), and emphasize the need for tools that measure internal vibrational coupling in molecules.

"…polariton chemistry might be used to perform molecular transformations that are intractable by conventional synthesis."

When molecules are strongly coupled to the cavity, each contributes a small vibrational amplitude to the collective polariton state in which all the molecules vibrate in sync. The remaining vibrational amplitude contributes to the so-called reservoir states, whose properties are not entirely understood. When all the molecules in solution are surrounded by similar microscopic environments, reservoir states involve unsynchronized vibrations of individual molecules, akin to molecules outside the cavity. However, when the environments slightly differ from one molecule to another, reservoir states can involve synchronized vibrations of multiple molecules (*11*). Chen *et al.* used laser pulses to excite $Fe(CO)_5$ in solvents, which creates similar environments for all the molecules in solution. The authors observed that vibrational dynamics were different only when the polariton state was excited, whereas upon excitation of the reservoir, the molecules behaved the same in the cavity as outside the cavity.

Chemists quantify the likelihood of reactions by the associated change in enthalpy and entropy. Synchronous vibrations of the polariton states presumably reduce the entropy of reactants with a potentially large impact on the reaction rate compared with relatively small changes in molecular vibrations caused by coupling to the cavity (*12*). However, in contrast to other reported reactions (*1*–*3*), Chen *et al.* did not observe substantial changes in the entropy and enthalpy of the pseudo-rotation and IVR upon coupling of $Fe(CO)_5$ to the cavity. It is not clear how these traditional chemical concepts can be used to interpret polariton chemistry.

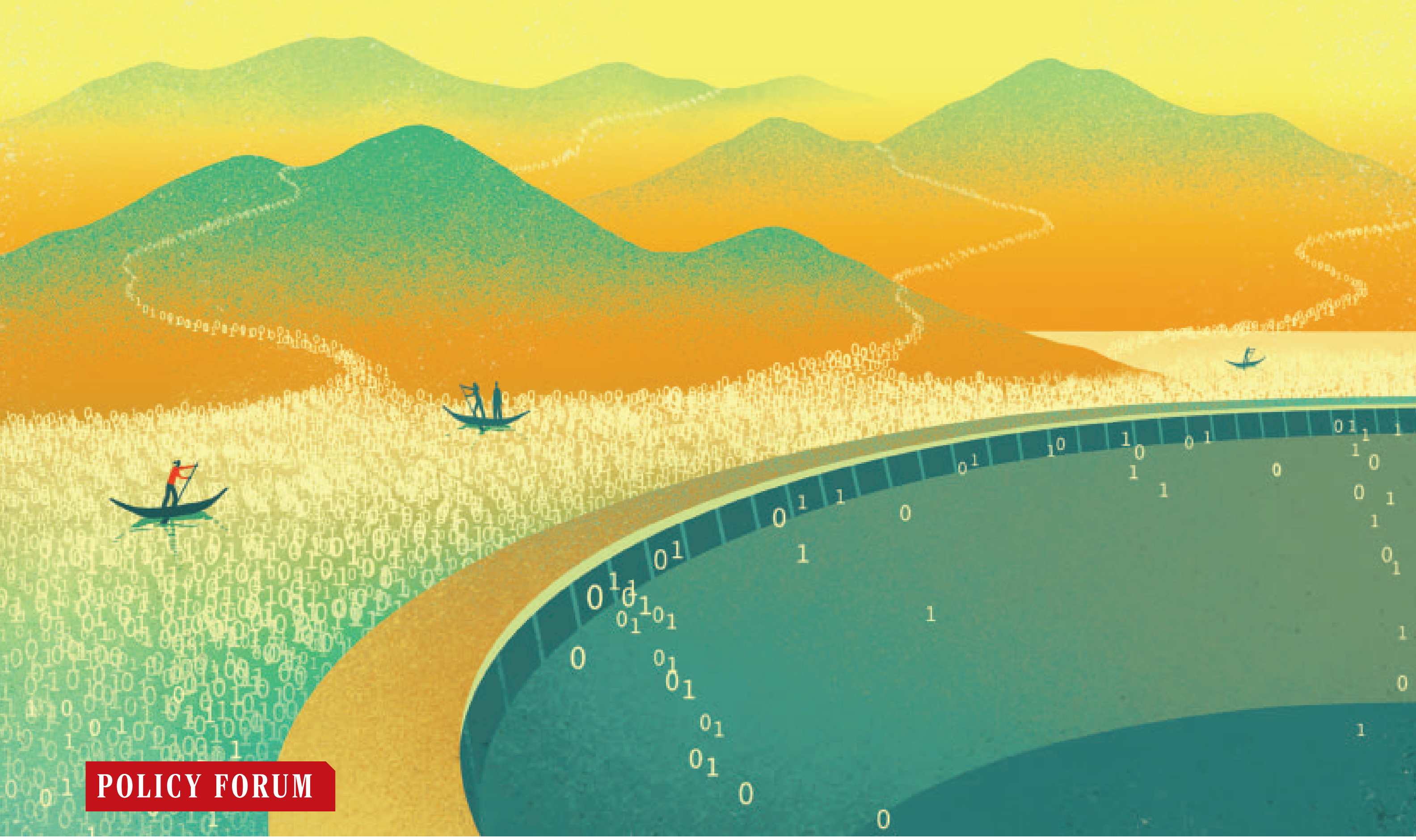
Chen *et al.* simultaneously excited molecules to the first excited vibrational state and from the first to the second excited state. To simplify the theoretical treatment, researchers frequently make approximations by considering only the two lowest-energy quantum states of the molecule. This leads to a theoretical model where the coupling strength between the molecules and the cavity depends on the number of molecules being excited (*13*). Such an approach, however, cannot describe experiments where dynamics of higher excited states must be considered (*14*). More fundamental research is needed to bridge between different interpretations of the experimental observations.

Polariton chemistry is in its infancy, and the fundamental principles driving uncharacterized phenomena are yet to be understood. The work of Chen *et al.* is a step toward a better understanding of such principles. ■

REFERENCES AND NOTES

1. A. Thomas *et al.*, *Science* **363**, 615 (2019).
2. J. Lather *et al.*, *Chem. Sci.* **13**, 195 (2022).
3. A. Sau *et al.*, *Angew. Chem. Int. Ed.* **60**, 5712 (2021).
4. B. S. Simpkins *et al.*, *J. Phys. Chem. C* **125**, 19081 (2021).
5. T.-T. Chen *et al.*, *Science* **378**, 790 (2022).
6. S. Karmakar, S. Keshavamurthy, *Phys. Chem. Chem. Phys.* **22**, 11139 (2020).
7. D. S. Wang *et al.*, *J. Phys. Chem. Lett.* **13**, 3317 (2022).
8. J. F. Cahoon *et al.*, *Science* **319**, 1820 (2008).
9. M. C. Thielges, M. D. Fayer, *Acc. Chem. Res.* **45**, 1866 (2012).
10. P. Portius *et al.*, *Organometallics* **38**, 4288 (2019).
11. B. Cohn *et al.*, *J. Phys. Chem. Lett.* **13**, 8369 (2022).
12. G. D. Scholes *et al.*, *J. Phys. Chem. Lett.* **11**, 6389 (2020).
13. R. Houdré *et al.*, *Phys. Rev. B* **52**, 7810 (1995).
14. C. A. DelPo *et al.*, *J. Phys. Chem. Lett.* **11**, 2667 (2020).

10.1126/science.ade9815



POLICY FORUM

DATA REGULATION

Research under China's personal information law

The new law may present obstacles to some kinds of research

By **Xiaojie Li**[1,2], **Yali Cong**[1,3], **Ruishuang Liu**[1,3]

The Personal Information Protection Law (PIPL) that came into effect in the Peoples Republic of China (PRC) in November 2021 is in line with many international standards because it was designed by referring to jurisdictions of other nations, especially the provisions of the European Union (EU) General Data Protection Regulation (GDPR). Thus, the fundamental principles of the PIPL—such as lawfulness, fairness, transparency, accuracy, and purpose limitation—are now broadly in line with the GDPR. However, China has not used "academic derogation" or "academic exemptions" as have been used in European nations, the United States, and Australia to reconcile the protection of personal data rights with scientific research freedom in data processing (*1*). We provide here a general overview of the effect of the PIPL on ongoing and future transnational scientific research through a legal analysis of its provisions, and we suggest feasible solutions for challenges brought by the PIPL to transnational scientific research.

The PIPL covers personal information of people in China, regardless of the location of the processing of information. In the research context, its provisions apply across the entire span of research processing, from the moment of data or biospecimens collection until the disclosure of research results. All personal information, except through anonymization processing that does not relate to an identified or identifiable person, is included in the regulation scope of the PIPL, no matter whether it was processed by scientific researchers, government agencies, companies, or others and no matter whether it was provided knowingly by people or collected without the knowledge of those whose information was collected (for example, by cameras in public spaces and cookies on web browsers).

We suggest a few potential explanations for why, unlike under GDPR, the PIPL does not adequately address the needs of contemporary scientific research. Most legislators believed that China was particularly insufficient in personal information protection, including in scientific research; thus, they felt that much stricter provisions were needed to protect personal information rights (*2*). Also, legislators did not fully consider the issues and implications around scientific research in the big data era, especially when the secondary use of personal information can be quite important. In part, this is because there was a relative lack of voices from the scientific community during the legislation process, a result of the legislation system in China.

With the law having taken effect only a year ago, there has not yet been a great deal of systematic monitoring and analysis of the implementation and impacts. Anecdotally, however, there is an impression that many ethical review boards and researchers are still in confusion or unclear about what is considered legal under the PIPL. Many researchers may

[1]Department of Medical Ethics and Law, Peking University Health Science Center, Beijing, China. [2]Department of Situation and Policy, University of International Business and Economics, Beijing, China. [3]National Institute of Health Data Science, Peking University, Beijing, China.
Email: lrs1187@vip.sina.com

ILLUSTRATION: DAVIDE BONAZZI/SALZMAN ART

still be using broad consent in their research and may not know how PIPL lacks space for this. Some researchers may be eager to consider ways by which broad consent for big data research may be made compatible with the PIPL.

PURPOSE LIMITATION

The PIPL requires that personal information shall be processed for explicit and reasonable purposes and be directly related to the purpose of processing in a manner that has a minimum impact on the rights and interests of individuals (article 6). If there is any change in the purpose or method of the processing of personal information, or the category of processed personal information, the individual is to be informed and their consent obtained again. The PIPL expects researchers to gain consent from the participant even if they have been deidentified, except by using "anonymization processing" (article 4). This severely limits the conditions of processing sensitive information without specific consent from the participant. Primary investigators (the processors) are required to seek voluntary and explicit consent from participants for the use of data or samples obtained from them by specified groups for a specific purpose that is necessary for the original study. For example, medical research institutions are typically not allowed to collect genetic information for research use unless they have a specific purpose (such as a multinational clinical trial) and sufficient necessity and take strict protective measures (article 28).

The process of scientific research and innovation in this era of big data is dynamic, full of uncertainties and possibilities. Determining all potential future specific purposes of data processing at the time of data collection is often not possible. If specific and explicit consent is required before the scientific research process begins, it may be almost impossible for many researchers to lawfully process the information. It may hinder advances in science and technology and the free flow of data. Furthermore, obtaining specific and explicit consent for secondary research often needs a disproportionate amount of effort, and finding the sources of data and specimens could be difficult or impossible. Such specification and identification of each type of processing without any derogation of personal data rights can substantially burden scientific researchers with additional bureaucratic work. A repeated informed-consent process that protects participants may dilute the purpose of research and what it can achieve while bringing unnecessary disturbances to participants (*3*). An exception from the purpose limitation principle may promote the use of data under many research scenarios.

SEPARATE CONSENT

Like the GDPR, the PIPL requires "specific consent," which means obtaining consent for a specific purpose, such as developing a drug for diabetes. But unlike the GDPR, PIPL also requires "separate consent," meaning that the authorization of multiple general types or features of processing (for example, processing of genetic information, processing of health care information, or change of information processors) cannot be bundled together to fall under a single umbrella authorization of individual consent. Instead, if there are multiple general types or features of sensitive personal information processing for which consent is sought, consent options should be set separately for each (*4*). The PIPL stipulates five scenarios in which obtaining separate consent is required, three of which are particularly relevant for transnational scientific research: (i) when processing sensitive personal information, such as biometric identification, health care, specific identification, financial accounts, religious beliefs, personal locations, and personal information of minors under the age of fourteen (articles 28 and 29); (ii) when personal information is provided to any processor outside the territory of the PRC (article 39); and (iii) when the information processors change (article 22).

"A repeated informed-consent process…may dilute the purpose of research and what it can achieve…"

Draft regulations on the management of online data security define separate consent as "the data processor obtains individual consent for each item of personal information when carrying out specific data processing activities, excluding the consent for multiple personal information and multiple processing activities at one time" (*5*). We believe that separate consent refers to obtaining personal consent for each type—such as biometric identification, financial accounts, and personal information of minors under the age of 14—rather than obtaining personal consent for each item. To define the scope of "separate," look to three points from article 23 of the PIPL: (i) The data processor has to perform specific data processing behavior or collect a specific type of data, (ii) the data processor has to inform each participant individually, and (iii) the process of obtaining informed consent has to be independent of other authorizations or should not be confounded with the provision of other services. If multiple types of information are used for which consent is sought, the corresponding types of information should be listed under each processing item, and consent should be obtained separately for each option.

SPECIMENS AND DATA TRANSPORT ACROSS BOUNDARIES

Compared with the relatively low limitation on free transfer of data and specimens under the GDPR, transport of data and specimens across boundaries in China is subject to stricter requirements under the PIPL. Personal information, reaching up to a specified quantity collected and generated within PRC territory by critical information infrastructure (CII) operators and personal information processors must be stored domestically (article 40) (for example, transport abroad of genetic information on 500 people or more needs administrative permission). If scientific researchers require transfer of personal information beyond the PRC, they must obtain separate consent from the participants and also meet one of the following conditions: (i) pass a security assessment organized by the Cyberspace Administration of China (CAC) in accordance with the provisions of article 40; (ii) be awarded personal information protection certification by a professional institution in accordance with provisions of the CAC; (iii) enter into a contract with the foreign recipient in accordance with the standard contract formulated by the CAC, stipulating the rights and obligations of both parties; or (iv) adhere to other conditions stipulated by laws, administrative regulations, or the CAC (article 38).

The presence of article 38(2) offers prospects for the free flow of data across boundaries under international treaties and agreements concluded or acceded to by China that stipulate the conditions for providing personal information outside China (for example, if China acceded to an agreement with other countries similar to the EU-US and Swiss-US Privacy Shield for commercial purposes). Furthermore, scientific researchers are to take necessary measures to ensure that the processing of personal information by foreign recipients meets the personal information protection standards stipulated in the PIPL under article 38(3). No specifications are available for determining whether the foreign receptions have met the PIPL's standard, which leads to uncertainty and requires the CAC to publish a standard contract for that purpose (*6*).

INDIVIDUAL RIGHTS AND DEROGATIONS FOR TRANSNATIONAL SCIENTIFIC RESEARCH

One key tenet of the PIPL, outlined in article 1, is that it prioritizes the rights of the data subjects (*2*). Accordingly, to ensure legiti-

macy and to comply with PIPL, researchers should ensure that subjects understand several rights when obtaining consent: the rights to be informed, to make decisions about the processing of their personal information, to access their information, to copy their information, to correct and supplement their information, as well as the right to erasure (to be forgotten), and the right to portability (to move data from A to B). Procedures for realizing these rights are elaborated, and requirements for processing sensitive personal information such as genetic data are stricter: "Only when there is a specific purpose, sufficient necessity and strict protection measures are taken, the processor can handle sensitive personal information" (article 28).

However, there are legal bases for a certain degree of derogation from the data subject's right to be informed and make decisions (article 44: otherwise provided by laws and administrative regulation), the right to delete (article 47: the circumstances specified in article 18, paragraph 1, and article 35), and the right to access and copy (article 45: if the retention period stipulated by laws and administrative regulations has not expired, or it is technically difficult to delete personal information), although these derogations are not as detailed as those in the GDPR.

LACK OF SPACE FOR BROAD CONSENT

Broad consent is an alternative to specific consent only in cases in which it is not feasible to obtain specific consent for scientific research (such as the situation in which the purpose of future research is uncertain and the secondary use of personal information is required in the future). Unlike blanket consent or general consent, broad consent must inform potential research subjects of certain matters concerning the long-term storage of biological samples or future studies with uncertain purposes: risks, benefits, confidentiality, right to refuse participation without reason, right to withdraw at any time, personal information processor information, whether biological samples are used for commercial benefit-sharing or whole-genome sequencing, a brief description of the field or type of research involved, storage duration, and research contacts and contact information. It is worth emphasizing that if specific consent is feasible after the purpose of future research is determined, then specific consent should still be preferred.

Unlike the GDPR, which includes provisions that seem to offer clear support for broad consent in scientific research, the PIPL requires specific consent for secondary use of personal information and specimens, which is inconsistent with research traditions and current ethical review guidelines. For example, the Ethical Review Measures for Biomedical Research Involving Human Subjects allow the donated sample and related information to be used for all kinds of medical research once the biological sample donor has signed the informed consent form (*7*). In the absence of restrictions, this kind of "blanket consent" allows samples and data to be used in future studies that may conflict with the individual's fundamental values. However, according to article 88 of the Legislation Law (*8*), this provision became invalid after the Regulations on the Administration of Human Genetic Resources (HGR regulation) and the PIPL took effect. Draft Ethical Review Measures for Life Sciences and Medical Research Involving Human Subjects clearly support broad consent by allowing secondary research to be carried out without reconsent under a wide range of circumstances (*9*). However, these provisions are contradictory to the PIPL, and the draft finally did not pass.

Under PIPL, scientific researchers may be granted exceptions from the requirement to collect consent when (i) performing legal duties or legal obligations; (ii) responding to public health emergencies or providing protection to the life, health, and property of individuals in emergencies; (iii) conducting news reports, public opinion supervision, and other acts for the public interest and processing personal information within a reasonable scope; (iv) processing personal information that has been disclosed by individuals themselves or that has been legally disclosed in accordance with the provisions of this law, within a reasonable scope; or (v) there are other circumstances stipulated by laws and administrative regulations (article 13). In the above scenario, the PIPL allows processing of personal information without consent but does not deprive individuals of their right to their personal information. But these scenarios would not likely apply to most transnational scientific research, which means that the PIPL lack space for broad consent. Fortunately, article 13(7) points out the direction of efforts in relation to scientific research: new laws and administrative regulations can be put forward in the future to support the legitimacy and utility of broad consent.

SUGGESTIONS FOR THE FUTURE

Although PIPL emphasizes personal information rights and interests, it does not fully consider the special circumstances of transnational scientific research in the big data era. It does not recognize how processing of personal data for research is different from other uses such as commercial purposes, neither do they notice that transnational scientific research can benefit society while balancing data subjects' rights (*10*). These concerns have been raised in the scientific research community.

To ensure that research is considered legal under the PIPL, we suggest some key steps. First and foremost, awareness of the requirements of PIPL must be increased, and the law enforced, in transnational scientific research efforts. Institutions such as the Peking University Health Science Center–Michigan Medicine (PKUHSC-MM) Joint Institute for Translational and Clinical Research should update informed consent forms and process personal information within the processing purpose and scope. Institutional review boards must be in place to ensure that the PIPL has been followed through training, guidance, and ethical review. Additionally, foreign researchers can promote the formulation of a treaty with China on transnational research or join a common international convention in accordance with article 38(2). Research experts should work toward consensus and provide proposals to the National People's Congress to characterize what constitutes scientific research and to legitimize exemption from portions of the PIPL for scientific research exemptions, such as the derogation of rights, the substitution of broad consent for specific consent under specific conditions, and the relaxation of purpose limitation. The advantages and disadvantages of applying PIPL should be evaluated as this law is being implemented. We believe that legislative adjustments should be made to ensure better balance among the rights and interests of data subjects, the wellness of the humankind, and the interests of researchers. ■

REFERENCES AND NOTES

1. M. Mourby *et al.*, *Int. Data Priv. Law* **9**, 192 (2019).
2. C. Xiao, Understanding and Application of Personal Information Protection Law (China Legal Publishing House, 2021).
3. D. Nicol *et al.*, *Science* **364**, 445 (2019).
4. X. Zhang, *Interpretation of "China Personal Information Protection Law"* (People's Press, (2021), p. 245.
5. CAC, "Regulations on the management of online data security (draft for comments)," November 2021; http://www.cac.gov.cn/2021-11/14/c_1638501991577898.htm.
6. A. Lee *et al.*, "Seven major changes in China's finalized personal information protection law" (Brookings, 2021); https://www.brookings.edu/articles/seven-major-changes-in-chinas-finalized-personal-information-protection-law.
7. The National Health Commission of the People's Republic of China (NHC), "The ethical review measures for biomedical research involving human beings (October 2016); http://www.nhc.gov.cn/cms-search/xxgk/getManuscriptXxgk.htm?id=84b33b81d8e747eaaf048f68b174f829.
8. The Ninth National People's Congress, The Legislation Law of the People's Republic of China (March,2000); http://www.gov.cn/xinwen/2015-03/18/content_2835648.htm.
9. NHC, The National Science and Technology Ethics Committee Measures for ethical review of life sciences and medical research involving human beings (March 2021); http://www.nhc.gov.cn/qjjys/s7946/202103/beb66b1525e64472b1a9b8921ed1aedf.shtml.
10. J. Bovenberg *et al.*, *Science* **370**, 40 (2020).

ACKNOWLEDGMENTS

All authors contributed equally to this work. Partial support came from the PKUHSC-MM Joint Institute for Translational and Clinical Research.

10.1126/science.abq7402



BOOKS *et al.*

HISTORY OF SCIENCE

The Gaia letters

Contextualized correspondence traces the emergence of a provocative hypothesis

By **Paul Falkowski**

Writing Gaia, edited by Bruce Clarke and Sébastien Dutreuil, is a fascinating read that reproduces and contextualizes a four-decade-long conversation between environmental scientist James Lovelock and evolutionary biologist Lynn Margulis from which emerged the provocative Gaia hypothesis, which posits that Earth and all its inhabitants can be thought of as a single, synergistic, self-regulating system.

The book opens with a foreword by the late Lovelock, in which he discusses a conversation he had in 1967 with William Golding—the Nobel Prize–winning author of *Lord of the Flies*—about Erwin Schrödinger's book *What Is Life?* "According to this book," observes Lovelock, "life was a process that reduced the entropy of a system while excreting entropy to the environment." Lovelock wanted to understand how Mars, with an atmosphere composed almost entirely of carbon dioxide, is of "high entropy" and therefore probably lifeless, whereas Earth, containing both methane and oxygen, is of "low entropy" and sustains life. "If you intend to put forward an idea like that," Golding replied, "you had better give the low-entropy system that is our planet a proper name, and I suggest the name Gaia."

The book goes on to document the four-decade-long correspondence between Lovelock and Margulis, which began in 1970, when Margulis, on the advice of her ex-husband Carl Sagan, first wrote to Lovelock, seeking his assistance in fleshing out research questions related to the contribution of biological entities to the planetary atmosphere. (That letter, unfortunately, is not included in this collection.)



Writing Gaia
Bruce Clarke and Sébastien Dutreuil, Eds.
Cambridge University Press, 2022.
510 pp.

The first letter in the book, dated 11 September 1970, from Lovelock states, "I am in the course of writing a paper on the Earth's atmosphere as a biological cybernetic system." The editors explain that by 1969, Lovelock had begun to think about Earth as a planetary ecosystem, and the term "biological cybernetic system" became a "shorthand trademark" for Gaia.

Although much of the correspondence is relatively trivial, detailing, for example, when and where they would meet, the editors did an exceptional job of culling the myriad letters and—with help from students and colleagues—interpreting the correspondence and revealing how the Gaia hypothesis evolved.

By the early 1970s, we learn, Margulis had become a major force for developing the idea that Earth systems function through feedbacks between life and the geosphere and gave Lovelock increasing confidence in the notion that "Gaia has the equivalent of a central nervous system." This notion would eventually lead to heated debates among many scientists, and the concept was, and still is, challenged, if not downright dismissed as an untestable hypothesis.

Arguably, one of the key questions surrounding the Gaia hypothesis is how original it was. The editors address this issue, discussing the contributions of Vladimir Vernadsky, a Soviet scientist who was one of the pioneers in a field that would become geobiology. Vernadsky published *Biosfera* in Russian in 1926, which was informally translated into English in the 1980s but not formally published in English until 1998.

James Lovelock and Lynn Margulis grappled with the possibility that Earth's life and its geosphere are intertwined in a single, synergistic system.

In the foreword of the 1998 edition, Margulis wrote, "[Vernadsky] illuminates the difference between an inanimate, mineralogical view of Earth's history, and an endlessly dynamic picture of Earth as the domain and product of life, to a degree not yet well understood." Lovelock disagreed; to him, Gaia was original. In June 1986, he wrote of Vernadsky to Margulis, saying, "he was a middle weight expressing his ideas in a vague and all-inclusive manner and with the support of little or no testable evidence."

Regardless of its originality, Lovelock persisted in developing Gaia as a theory and, to his death, defended the notion of a positive feedback between the biosphere and geosphere. He and Margulis thought the concept was as seminal as Darwin's theory of evolution. Indeed, in 1995, Lovelock wrote, "Where organisms affected their personal environment then the tendency could be inherited and could become extensive, even global."

At almost 500 pages, *Writing Gaia* is a weighty read. The more casual reader can skim the correspondence and read the editors' excellent summaries. Others will find the details of the correspondence a fascinating illustration of how ideas between scientists evolve. I am sure this tome will be used by many students and scholars to understand the history of geobiology, writ large, in the 20th century. ■

10.1126/science.ade8916

The reviewer is at the Department of Earth and Planetary Sciences and the Department of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ, USA. Email: falko@marine.rutgers.edu

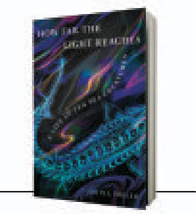
PHOTO: TUI DE ROY/MINDEN PICTURES

SCIENCE AND SOCIETY

Aquatic life and its parallels

Blurring the personal and the scientific, an author probes the beauty, terror, and richness of the natural world

How Far the Light Reaches: A Life in Ten Sea Creatures
Sabrina Imbler
Little, Brown, 2022.
320 pp.



By **Juli Berwald**

In 1996, *Scientific American* staff writer John Horgan published *The End of Science*, which posited that all the great discoveries had been made and science was reaching its twilight (*1*). But from within those pages, a loophole emerged. In one chapter, physicist David Bohm pointed out that art and science, now separate pursuits, have not always been so. During the Renaissance, for example, the two intermingled, and great progress was made.

But as science became more objective, it also became more erudite and less relatable. Facts, although powerful, eventually cave under the weight of emotion. Vaccine hesitancy and climate denial, to name two big examples, suggest the limits of science, which is constrained not by all possible discoveries having been made but by the fact that humans tend to comprehend ideas yet act on feelings.

What would happen if art, whose main purpose is to make us feel, again merged with science? Writer and science journalist Sabrina Imbler's second book, *How Far the Light Reaches*, points toward an answer. The book's 10 essays, each musing on a different aquatic creature, upend paradigms of culture and history and force readers to confront their assumptions about society.

Imbler's pieces often start in one place and end up somewhere else entirely. "Beware the Sand Striker" describes what it means to be preyed upon, bringing together painful stories of Imbler's sexual encounters with men and the whip-fast strike of a polychaete worm that hides in the sandy seafloor, invisible to the fish and crabs above. It also tells the true story of Lorena Bobbitt, whose married surname has been exploited as a nickname for the same worm.

In a chapter called "Hybrids," Imbler explores "The Question" of being biracial, intertwining it with the story of a hybrid butterflyfish collected in Australia in the 1970s, in the process turning a searing eye on a celebrated ichthyologist. In "How to Draw a Sperm Whale," Imbler describes whale necropsy—the animal equivalent of an autopsy—and then uses it to dissect the death of a relationship with a lover.

As a writer, Imbler is a terrific talent, at once commanding and questioning, unflinching and vulnerable, factual and poetic, polished and raw, specific and general. The book opens with the words "The truth is," which the reader should take seriously. Imbler bares their soul and leaves no room for platitudes.

Some of the book's most powerful essays revolve around Imbler's growing understanding and recognition of their queerness. They contemplate the biology of a goldfish, confined to a bowl growing ever more polluted. The same chapter tells of Imbler's struggle as an overachieving rule-follower in high school. And then, years later, their first gay encounter with someone unexpectedly from their hometown. "Release a goldfish, and it will never look back," Imbler writes. "Nothing fully lives in a bowl; it only learns to survive it."

In another essay, contemplating the seemingly boundless morphing of a cuttlefish's color and texture to avoid predation, Imbler asks, "But once the danger passes, you may feel tempted to seize this metamorphic power for means beyond escape. When you are not fleeing, what will you become?"

Part of a solution is discovered in the essay "Pure Life," which looks into the deep sea where, along hydrothermal vents, the yeti crab dances with its claws in the air, bringing sulfur-laden water to the chemosynthetic bacteria that sustain it. In the dark nights and gay bars of Seattle, Imbler, too, learns what it means to dance to be alive, finally among a community of people who make them feel seen and known.

The theme that pulls these essays together is that the distinctive characteristics that make the author who they are, which are often framed as outliers and differences, are all reflected throughout the natural world. Nature's greatest gift is its creativity and diversity.



Like a goldfish in a bowl, we can survive in stifling environments, but we thrive when our needs are fully met.

PHOTO: MATT LEE/ALAMY STOCK PHOTO

But there is something else going on in Imbler's work. Like Horgan did in 1996, Imbler may make readers uncomfortable. Not because they portend a dismal future, but because, with brutal candor and elegant metaphor, *How Far the Light Reaches* reveals the gap between where we are today and a truly inclusive and connected world. In so doing, it also threads the loophole, weaving the outlines of a future where art and science amplify one other. ■

The reviewer is the author of *Life on the Rocks: Building a Future for Coral Reefs* (Riverhead Books, 2022). Contact: www.juliberwald.com

REFERENCES AND NOTES

1. J. Horgan, *The End of Science: Facing the Limits of Knowledge in the Twilight of the Scientific Age* (Broadway Books, 1996).

10.1126/science.ade9267

LETTERS

Edited by **Jennifer Sills**

Editorial Retraction

On 21 July 2017, *Science* published the Report "Chiral Majorana fermion modes in a quantum anomalous Hall insulator–superconductor structure" by Q. L. He *et al.* (*1*). Readers who failed to reproduce the findings requested raw data files from the authors, which they provided. Subsequently, the provenance of the raw data came into question; additionally, an analysis of the raw and published data revealed serious irregularities and discrepancies. These issues have caused the editors at *Science* to lose all confidence in the conclusions of the paper, and we are therefore proceeding with an Editorial Retraction. Authors A. L. Stern, J. Wang, and B. Lian agree with this decision. Authors Q. L. He, L. Pan, X. Che, G. Yin, E. S. Choi, K. Murata, X. Kou, Z. Chen, T. Nie, Q. Shao, Y. Fan, K. Liu, J. Xia, and K. L. Wang disagree with this decision. Authors E. C. Burks and Q. Zhou did not respond. Author S.-C. Zhang is deceased.

This Retraction replaces the Editorial Expression of Concern posted on 16 December 2021 (*2*).

H. Holden Thorp
Editor-in-Chief

REFERENCES AND NOTES

1. Q. L. He *et al.*, *Science* **357**, 294 (2017).
2. H. H. Thorp, *Science* **374**, 1454 (2021).

10.1126/science.adf7575

Brazil's preventable bridge disasters

In October, a second bridge collapsed within 10 days on Brazil's controversial BR 319 highway (*1*), as biologists predicted could happen (*2*). Bridges collapse regularly in Brazil, and the events are often associated with flooding (*3*). Brazil's environmental legislation theoretically protects locations that are likely to flood from development, but a legal loophole allows infrastructure projects to proceed if they are considered to be in the public interest. If Brazil's government does not take action to restrict use of the loophole, preventable disasters will continue to occur.

Brazilian Law 12,651, known as the Forest Code (*4*), was enacted to prevent natural disasters and protect infrastructure. The legislation designates Permanent Protection Areas around water bodies to guarantee the health of riparian areas that can buffer floods. According to the law, construction is forbidden in these important ecosystems. The law also protects riparian systems in the Amazon, which has heavy rainfall, fragile soils, and often superficial water tables.

Earth-moving companies are the primary beneficiaries of the law's loophole. The companies argue that the cost savings of bulldozing earthen access ramps across protected riparian areas justifies the added risks. However, the embankments they build funnel the main flow of flood waters around the pylons of bridges, increasing the pressure and the likelihood of collapse. In some cases, the bridge holds up and only the access ramps are washed away, after which the earth-moving companies are paid to restore the road. The companies benefit, perhaps unwittingly, but the economy and local population suffer.

The Forest Code is good, evidence-based policy, but it must be enforced wisely to maximize its impact. The loophole was included so that infrastructure truly in the public good, such as bridges required for access, could be constructed in the riparian zone. The policy was not designed to license the wholesale destruction of protected areas, along with their essential ecosystem functions in flood mitigation. Brazil must acknowledge that it is almost never in the public interest to destroy Permanent Protection Areas and risk provoking disasters.

William E. Magnusson
Instituto Nacional de Pesquisas da Amazônia, Manaus, Amazonas, Brazil.
Email: wemagnusson@inpa.gov.br

REFERENCES AND NOTES

1. H. Cristaldo, "Amazonas registra segundo desabamento de ponte na BR-319," *Agência Brasil* (2022); https://agenciabrasil.ebc.com.br/geral/noticia/2022-10/amazonas-registra-segundo-desabamento-de-ponte-na-br-319 [in Portuguese].
2. C. Rosa *et al.*, *Aquat. Conserv. Mar. Freshw. Ecosys*. **31**, 1548 (2021).
3. "Queda de pontes e viadutos no Brasil," *Folha de São Paulo* (2018); https://fotografia.folha.uol.com.br/galerias/1617195713692955-queda-de-pontes-e-viadutos [in Portuguese].
4. Câmara dos Deputados, "Legislação Informatizado – Lei n° 12.651, de 25 de maio de 2012 – publicação original" (2012); https://www2.camara.leg.br/legin/fed/lei/2012/lei-12651-25-maio-2012-613076-publicacaooriginal-136199-pl.html [in Portuguese].

10.1126/science.adf2868

Welcoming Taiwan's diaspora scientists

Academia Sinica, the highest research academy in Taiwan, recently announced 19 newly elected Members and 3 Honorary Members (*1*) after a 2-year pandemic delay. Because in 2020, the Taiwanese authority declared that only Taiwanese citizens could be nominated as Members of Academia Sinica (*2*), another five candidates remain



A loophole has allowed construction in Brazil's protected areas, contributing to multiple bridge collapses.

PHOTO: BRUNO KELLY/REUTERS

pending (*1*). This decision undermines the inclusiveness of Academia Sinica and the scientific community in Taiwan.

Many small countries and territories—including Taiwan—have suffered from "brain drain" (*3*) as scientists leave to study and work elsewhere. Elections of diaspora scientists into the academies of their home countries and territories provide an opportunity to keep them connected to their homelands, in some cases even influencing their decision to return (*4*–*6*). The Hungarian Academy of Sciences has External Memberships for diaspora Hungarians, as well as Honorary Memberships for non-Hungarian foreigners. In Colombia (*7*), Israel (*8*), mainland China (*9*), and African countries (*10*), efforts to connect with diaspora scientists include research scholarships and cooperation initiatives that facilitate positive interactions between overseas scientists and scientific developments in their native countries.

A small homeland like Taiwan can benefit from enhancing the local influence of scientists who have left (*11*). The new citizenship requirement for member elections at Academia Sinica in Taiwan could instead decrease the engagement of scientists abroad (*12*). Given the challenges Taiwan faces, the Taiwanese authority should adapt its policy to be more welcoming of its diaspora scientists. In addition to piloting an External Member system, similar to those in Hungary and Columbia, Taiwan should counter brain drain by increasing the opportunities available to overseas scientists and adopting more flexible citizenship requirements.

Jianan Huang
Nanyang Centre for Public Administration, Nanyang Technological University, 639798 Singapore and National Institute of Education, Nanyang Technological University, 639798 Singapore. Email: jhuang067@e.ntu.edu.sg

REFERENCES AND NOTES

1. Academia Sinica, *Academia Sinica Newsletter* (*Eng. Ver.*) 1769 (2022).
2. Legislative Yuan of ROC, *Legislative Yuan Bulletin* **72**, 109 (2020) [in Chinese].
3. M. U. M. Anas, S. I. Wickremasinghe, *Sci. Public Pol.* **5**, 37 (2010).
4. Y. Wang *et al.*, *Bull. Chin. Acad. Sci. (Chin. Ver.)* **1**, 34 (2019) [in Chinese].
5. C. Cao, *China's Scientific Elite* (Routledge, 2004).
6. C. Cao, *Science* **278**, 785 (1997).
7. J.-B. Meyer *et al.*, *Sci. Technol. Soc.* **2**, 285 (1997).
8. O. Rabinowitz, Y. Abramson, *Soc. Stud. Sci.* **2**, 52 (2022).
9. D. Zweig, S. F. Chung, D. Han, *Sci. Technol. Soc.* **1**, 13 (2008).
10. H. Pellerin, B. Mullings, *Rev. Int. Polit. Econ.* **20**, 89 (2013).
11. Y. Shain, A. Barth, *Int. Org.* **57**, 449 (2003).
12. Legislative Yuan of ROC, *Legislative Yuan Bulletin* **83**, 111 (2022) [in Chinese].

10.1126/science.adf3523


The Asian elephant's anatomy could hold the key to questions in a variety of scientific fields.

PHOTO: WPHIL MCLEAN/FLPA/SCIENCE SOURCE

LIFE IN SCIENCE

The elephant in the inbox

"We'll need to get you some boots," the veterinarian said to me, pointing toward my clean, shiny dress shoes. "Excuse me?" I replied naively, beginning to regret my decision to check my emails at 5 p.m. on a Friday. I had come to work dressed for an after-work dinner date, not for what I was about to experience.

An hour before, I had been shutting down the microscope and finishing up what had been a normal week as a postdoc in a muscular metabolism lab here in Copenhagen. Looking forward to what I thought would be a romantic evening, I scanned my inbox one last time. Seven new emails caught my attention, all sharing a title: Re: Elephant.

Copenhagen Zoo houses more than 3000 animals, and despite receiving excellent care, one of them occasionally requires euthanasia. The tragic death of such a unique animal can provide a valuable opportunity for scientists. The Asian elephant (*Elephas maximus*) is one of the largest existing land mammals, and powerful animals require powerful muscles. By studying how an elephant's muscles function, we may be able to better understand how our own muscles work.

In my career, I have worked with mice, rats, and worms. Yet when I clicked on the emails, I found a request to cycle to the local zoo immediately with some ice and collect a sample of elephant muscle from the front gate.

So there I was, in my best shoes, clutching my box of ice and suddenly realizing why boots were going to be required. There would be no quick handoff at the zoo entrance as the emails had implied. Over the next 4 hours, I assisted a team of veterinarians in the postmortem of this gigantic animal, as a forklift and heavy machinery harvested organs to send to teams of researchers across Europe. I learned that tissue would be used in research for vaccines, cardiac disease, diabetes, and infectious disease, as well as for my lab's research into muscular weakness.

It was dark by the time I stored the tissue. Five hours late for my dinner, I left the boots behind and went to beg forgiveness. Fortunately, my date forgave me. We now live together, and we have a no-email rule for Friday nights.

Call for submissions
Life in Science is an occasional feature highlighting some of the humorous or unusual day-to-day realities that face our readers. Can you top this? Submit your story to www.submit2science.org.

Christopher T. A. Lewis
Xlab, Center for Healthy Aging, Department of Biomedical Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.

10.1126/science.adf5850



CREDIT: (PHOTO) MIT MUSEUM/CC BY-NC; (PAGE ELEMENTS) ADAPTED FROM KYOSHINO/ISTOCK.COM AND SAMGRANDY/ISTOCK.COM

PERSPECTIVES

Moore's law: The journey ahead p. 722

Toward gallium oxide power electronics p. 724

REVIEWS

Carbon nanotube transistors: Making electronics from molecules p. 726

Toward attojoule switching energy in logic transistors p. 733

RELATED ITEM

EDITORIAL p. 683

FROM ONE TRANSISTOR...

By **Phil Szuromi**

For most of 1947, the count of transistors made was...zero. It is now estimated that at least 3 sextillion transistors have been made and shipped since then. What started as a fundamental experiment probing the nature of electronic structure at semiconductor surfaces ended up containing a seed for applied science—it was soon recognized that the three-electrode device could amplify signals and replace vacuum-tube amplifiers. The development of integrated circuits spawned the ongoing race to continually increase the number of transistors on a chip and lower the cost and energy use per transistor. This special issue recaps some of this history and offers glimpses into how transistor technology may move forward.

The reach of transistors is difficult to overestimate. Transistors not only transformed communications but also made computers widely available. In science and technology, high-power electronics enable gel electrophoresis of DNA and extensive computational and software resources enable genomic sequencing. The transistor-based Apollo mission guidance computers executed complex control tasks in real time, with the help of reliable software to operate them.

The discovery of transistors speaks to the importance of asking fundamental questions and being aware of potential applications that could be hidden in the answers. The development of transistors into modern integrated circuits over the past 75 years was the result of the efforts of huge teams of researchers and engineers. If used wisely by the rest of us, transistors can continue to extend our scientific and technological reach.

Margaret Hamilton presents the listings of the software she and her team developed for the transistor-based Apollo guidance computer.

DEVICE TECHNOLOGY

Moore's law: The journey ahead

High-performance electronics will focus on increasing the rate of computation

By **Mark S. Lundstrom** *and* **Muhammad A. Alam**

The transistor was invented 75 years ago, and the integrated circuit (IC) soon thereafter. The progress in making transistors smaller also led to them becoming cheaper, which was famously noted as Moore's law (*1*). Today's sophisticated processor chips contain more than 100 billion transistors, but the pace of downsizing ("scaling") has slowed and it is no longer the only or even main design goal for improving performance in particular applications. How can Moore's law continue on a path forward? New approaches include three-dimensional (3D) integration that will focus on increasing the rate of information processing, rather than on increasing the density of transistors on a chip.

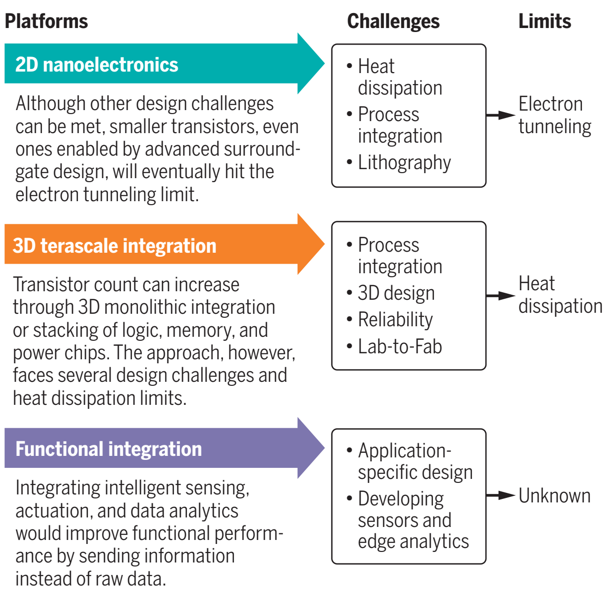
Although Moore's law predicted a rate for the decrease in cost per transistor, it is popularly viewed in terms of transistor size, which for two-dimensional (2D) chip arrays translates into an areal size or "footprint." During the last 75 years, as the footprint has decreased from micrometer to nanometer scales, issues with implementing new fabrication technologies have raised concerns several times about the "end of Moore's Law." Twenty years ago, a pessimistic outlook prevailed regarding the development of several difficult technologies for scaling to continue. In this context, one of the authors (M.S.L.) predicted that instead of slowing down, the scaling of metal-oxide-semiconductor field-effect transistors (MOSFETs) below the so-called 65-nm node, which was state-of-the-art in 2003, would continue unabated for at least a decade before the scaling limit was reached (*2*).

Scaling indeed continued from about 100 million transistors per chip in 2003 to as many as 100 billion transistors per chip today. One approach was to improve the on-off current ratio to allow practical operation and suppress leakage current to reduce wasted power. In 2003, strained silicon was introduced as channel material, and it increased the on-current by increasing the velocity of electrons (*3*), and in 2004, gate insulators with a high dielectric constant decreased the off-state gate-leakage current. In 2011, the FinFET, a nonplanar transistor structure that increases the electrostatic control of the energy barrier by the gate electrode (and thereby improves the on-off current ratio), was introduced in commercial ICs. Gate all-around transistors that further improve the electrostatic control of the gate are now in development (*4*). The size of transistors that can be fabricated is limited by patterning and etching. Patterning is done by a process known as photolithography, in which a photoreactive polymer creates a mask on the chip for etching steps. The minimum size of the pattern is determined by the wavelength of the light used. The recent emergence of extreme ultraviolet lithography (EUV) made it possible for Moore's law to continue beyond the 7-nm node (*5*).

School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907-1971, USA. Email: lundstro@purdue.edu

Three platforms forward

Two-dimensional (2D) nanoelectonics, three-dimensional (3D) terascale integration, and functional integration can all extend Moore's law, but all face substantial challenges and fundamental limits.



GRAPHIC: K. HOLOSKI/*SCIENCE*

The number of transistors on a chip is still increasing, but the rate of scaling has slowed because smaller transistors do not function very well. Specifically, the length of the channel (the region between the source and drain electrode where the gate acts as a switch) is now ~10 nm. At shorter channel lengths, excessive quantum-mechanical tunneling degrades transistor action. Key performance metrics, such as on-current (which should be high for high-speed operation), off-current (which should be low to minimize standby power), and power supply voltage (which should be low to minimize the power consumed), all degrade simultaneously. Silicon MOSFETS are now about as small as they can get, and the 2D chips are about as large as they can be made, so new ways to advance performance must be found.

Performance is being enhanced by moving from general-purpose, "commodity chips" to ones that accelerate specific functions. For example, hardware acceleration offloads specific tasks to specialized chips such as graphics processing units or an application-specific IC. Companies such as Apple now design their own chips to meet their specific requirements, as will all of the major automobile manufacturers. Computing is the limiting factor for machine learning, and companies such as Google now design their own artificial intelligence (AI) accelerator chips. Custom chip designs can increase performance by orders of magnitude, but just as the cost of chip manufacturing facilities ("fabs") has multiplied (from ~$1 billion in 2000 to ~ $20 billion for a leading-edge fab), so has the cost of leading-edge design. The design of a leading-edge chip can cost $0.5 billion and require a team of ~1000 engineers. Lowering the cost of leading-edge, custom-chip design (possibly by using machine learning techniques) will be a key challenge for the next era of electronics.

Continued progress will also require advances in the underlying technology. Despite the sharp increase in the number of transistors on chips (both by decreasing

their size and increasing the 2D chip area), until recently one aspect of the design had remained largely unchanged. Individual chips are packaged and combined with other chips and other components (such as inductors) laterally on printed circuit boards. Sending signals on- and off-chip increases delays and power consumption.

An emerging design theme is to exploit the third (vertical) dimension to enable terascale integration (TSI), with trillions of transistors integrated into monolithic or stacked chips and terabits per second per millimeter communication speed for electrical or optical interconnections ("per millimeter" refers to the communication link distance between the chips). For example, a 3D NAND flash memory (which is based on NAND logic gates and retains its states with the power off) can have nearly 200 layers of devices and half a trillion memory transistors (*6*). Emerging logic transistors with new channel materials (such as transition metal dichalcogenides and indium oxide) that can be processed at low temperatures and embedded within the interconnect stacks offer further opportunities.

The third dimension also opens up the possibility of vertical heterogeneous integration of logic, memory, and power transistors. With "through-Si vias," metal wires that connect vertically from the chip, already-processed chips can be stacked to put them in close physical proximity to minimize signal delays and reduce power consumption (*7*). Vertically stacked logic and memory chips also enable new computing paradigms, such as "compute-in-memory." Monolithic 3D ICs would consist of layers of active devices, such as 2D logic transistors, magnetoresistive and resistive random-access memories, and ferroelectric FETs, along with the metal lines that interconnect them (*8*).

Recent packaging innovations, such as silicon-interposer and multi-die silicon bridges, inserted between the 3D chips and the substrate, create denser lateral interconnection and faster communication among the chips. Advanced packaging brings together logic, memory, power management, communications, and photonics through side-by-side integration. The proximity of integration now rivals that in stacked and monolithic 3D ICs (*8, 9*).

Monolithic 3D integration will require that the growth or deposition steps do not affect the already-processed layers. For example, the transistors embedded within the interconnect stack must be deposited at a low-enough temperature not to disturb the dopant profiles of the Si transistors underneath. The needed materials are often incompatible unless special processes are developed. Stacking already-processed 2D chips to achieve 3D systems has its own set of material and processing challenges, such as maintaining interconnect alignment over distances of ~1 to 5 μm. Heterogeneous integration of components such as Si high- and low-voltage logic and memory transistors, and compound semiconductor-based power and high-frequency transistors, presents another set of complex integration challenges.

Transistors generate heat when they operate, and removing the heat is a serious issue in electronics today (*7, 10*). Indeed, thermal cross-talk among logic, memory, power transistors, and inductors in a heterogeneous IC creates unprecedented design challenges. New ways to remove heat, perhaps mimicking the thermoregulation of organisms, and thermally aware design will be critical when trillions of transistors are placed in close proximity.

The reliability of electronic systems must be guaranteed for a minimum time, typically 10 years, but decades for some applications. Ensuring a failure rate between 1 and 10 parts per million for ICs with 100 billion transistors each requires predicting the reliability of quintillion (~10^{18}) transistors. In practice, reliability is determined through short-term accelerated testing of no more than a few thousand transistors. Thus, the reliability physics of the wear-out and catastrophic failure modes of these new systems need to be understood with unprecedented precision. When so many devices are interconnected and placed in close proximity, new phenomena will emerge, and these must be managed or exploited.

Future terascale systems will be fundamentally different from today's gigascale systems in that understanding the building blocks of a system do not inform how these blocks interact and lead to new phenomena (*11*). Chip design is already complex and expensive, but algorithms or tools to place devices for 3D design and routing the interconnections among them are not yet available. These design tools must model the complexity of the process and package integration, thermal cross-talk among 3D ICs, and operation-specific variability and reliability of the packaged system.

When new materials and processing techniques are developed in research, they must be translated to large-scale manufacturing. Translating advances achieved with research-grade equipment to large-scale manufacturing with different and more sophisticated state-of-the-art manufacturing equipment presents a serious "lab to fab" challenge. The research community will need access to advanced processing facilities, and short "conceive-conduct-analyze" experimental loops that maximize learning are needed.

Thermal issues will define the limits of 3D terascale integration, just as tunneling limits have stymied 2D scaling. This requirement need not herald an end of Moore's law. The goal of computing is not operations per second but information per second. In that regard, biology offers a guide. Human senses process information locally before forwarding it to the brain. Empowering the sensors at the edge that interfaces the analog world, supported by local memory and data processing (edge analytics), could prevent the data deluge from overwhelming the computer.

Electronics is at an inflection point. For 75 years, it has been possible to make transistors smaller, but that will not be the driving force for progress in the decades ahead. If Moore's law is understood to refer to the increasing number of transistors per integrated system (not necessarily per chip), then the end of Moore's law is not in sight (see the figure). The increasing number of transistors will not come by making them smaller, but by stacking them vertically or combining them laterally in sophisticated packages, and eventually in monolithic 3D chips and adding functionality.

Shifting from nanoelectronics (focused on reducing the transistor dimension) to terascale electronics (driven by increasing transistor count and related functionality) defines the paradigm shift and core research challenges of the future. It will require fundamental advances in materials, devices, processing, and the design and manufacture of the most complex systems humans have ever built. Someday the electrical tunneling and thermal bottleneck will define the limits of 3D integration. Until then, Moore's law will likely continue as researchers address the challenges of these extraordinarily complex electronic systems. ■

REFERENCES AND NOTES

1. G. E. Moore, Electronics (Basel) no. 8 (19 April 1965) (1965).
2. M. Lundstrom, *Science* **299**, 210 (2003).
3. K. Mistry *et al.*, in *IEEE International Electron Devices Meeting*, pp. 247250 (2007).
4. D. Jang *et al.*, *IEEE Trans. Electron Dev*. **64**, 2707 (2017).
5. WIRED, "The $150 million machine keeping Moore's law alive," 30 August 2021; https://www.wired.com/story/asml-extreme-ultraviolet-lithography-chips-moores-law/.
6. A. Goda, *Electronics* **10**, 3156 (2021).
7. R. W. Keyes, *Proc. IEEE* **60**, 225 (1972).
8. S. Iyer, *IEEE Trans. Compon. Packaging Manuf. Technol*. **6**, 973 (2016).
9. C. H. Douglas *et al.*, in 2021 *IEEE International Electron Devices Meeting*, pp. 3–7.
10. M. A. Alam *et al.*, *IEEE Trans. Electron Dev*. **66**, 4556 (2019).
11. P. W. Anderson, *Science* **177**, 393 (1972).

10.1126/science.ade2191

DEVICE TECHNOLOGY

Toward gallium oxide power electronics

Ultrawide-bandgap semiconductors show promise for high-power transistors

By **Marko J. Tadjer**

Efficient, ultrahigh-voltage power-conversion electronics (with voltages >20 kV) require semiconductors with an energy gap much larger than that of silicon. The wide-bandgap (WBG) semiconductor silicon carbide (SiC) has matured into a commercial technological platform for power electronics (*1*), but ultrawide-bandgap (UWBG) (bandgaps >4.5 eV) semiconductor devices could potentially enable substantially higher-voltage electronics. Candidate UWBG semiconductors include aluminum nitride (AlN), cubic boron nitride, and diamond, but during the past decade, the greatest increase in research activity has likely been directed at gallium oxide (Ga_2O_3). This interest is driven in part by its large bandgap of ~4.85 eV and breakthroughs in crystal growth that led to the first Ga_2O_3 transistor demonstration in 2012 (*2*). Ga_2O_3 has promise as a platform for power electronics, but there are challenges in bringing this UWBG semiconductor into commercial use within the next decade.

The electrification process that has captured the attention of many industries could be disruptively accelerated if ultrahigh-voltage electronics penetrate into applications such as next-generation power-grid control and protection, ultrafast electric vehicle chargers, or efficient point-of-load converters with size, weight, and power advantages. Although SiC devices are higher in cost compared with conventional silicon power electronics, at the system level, those costs are expected to be offset by savings given simpler circuitry requirements.

Power conversion at very high voltages and high switching speeds may become attainable beyond 20 kV if a viable UWBG technological platform emerges. Even at 10 kV, it is difficult to increase the switching frequency of a power converter beyond ~10 kHz without sacrificing circuit efficiency. UWBG semiconductors inherently require thinner device layers that lead to reduced conduction losses (proportional to channel resistance). The reduced carrier transit time through a smaller UWBG device would also reduce switching losses (proportional to capacitance) and provide a platform for higher-speed electronics without sacrificing output power. Such high-speed power transistors would be disruptive in the power electronics industry because system volume is inversely proportional to frequency.

Out of the six crystalline Ga_2O_3 phases, the low-symmetry monoclinic β-Ga_2O_3 is furthest along in its development cycle because of its thermal stability at high temperatures (>650°C) (*3*), and the discussion below refers to this phase. Unlike any other WBG or UWBG semiconductor, melt growth methods originally developed for silicon substrates have been adapted to commercialize Ga_2O_3 substrates (*4*). β-Ga_2O_3 wafers have reached the 4-inch (100-mm) commercial milestone and are on track to become available at the 6-inch (150-mm) size by 2027 (*5*). In parallel, the infrastructure for high-quality epitaxy is being scaled to keep up with the increasing Ga_2O_3 substrate size. Methods for Ga_2O_3 epitaxial growth, such as chemical vapor deposition (CVD), molecular beam epitaxy, and halide vapor phase epitaxy among others, are being extensively researched with the goal of producing the highest-quality material.

Although the basic infrastructure building blocks for UWBG technology are well into their development cycle, UWBG device architectures are still actively explored by researchers. A vertical field-effect transistor (FET), such as the FinFET (see the figure, left), can in theory block very high fields but is more susceptible to extended defects in the epitaxial layer. A lateral transistor, such as the heterojunction FET (see the figure, right), could potentially switch faster and more efficiently because of its smaller capacitances and shorter transit times, and it could also use ternary alloys of Ga_2O_3, in this case β-$(Al_xGa_{1-x})_2O_3$, to boost power performance even further.

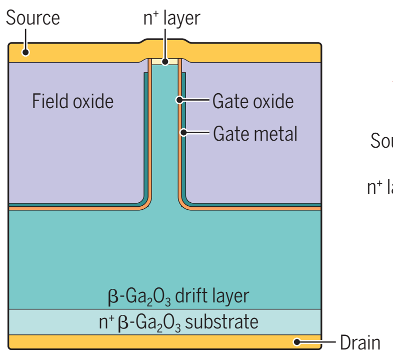
The existence of shallow energy donors and acceptors (charged impurities) has plagued all UWBG semiconductors because an increasingly wider energy gap typically leaves an extrinsic impurity to reside farther from the conduction (or valence) band. However, for Ga_2O_3, silicon is an excellent extrinsic shallow donor that has enabled a very wide range of controlled conductivities, from below 10^{14} cm^{-3} to above 10^{20} cm^{-3}. The controllable n-type conductivity extends even to the ternary alloy $(Al_xGa_{1-x})_2O_3$, which has an even wider bandgap that is also tunable depending on phase and Al concentration (*6*). Also, CVD-grown homoepitaxial β-Ga_2O_3 has a purity surpassed only by silicon. Recently, extremely high low-temperature mobility

Gallium oxide (β-Ga_2O_3) devices

The availability of wafer substrates and growth processes enables fabrication of devices for power electronics.

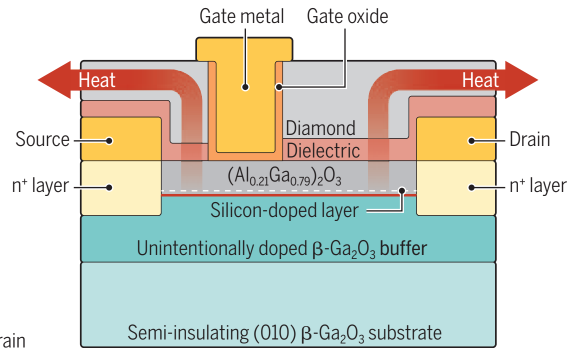
High-field operation

Schematic of a vertical β-Ga_2O_3 fin field-effect transistor (FinFET) with 4.2-kV breakdown voltage.



Thermal management

Schematic of near-junction integration of nanocrystalline diamond with β-Ga_2O_3 transistor for top-side heat extraction.



GRAPHIC: A. FISHER/*SCIENCE* BASED ON NOVEL CRYSTAL TECHNOLOGY, INC.

United States Naval Research Laboratory, Washington, DC 20375, USA. Email: marko.tadjer@nrl.navy.mil

(23,000 cm^2 V^{-1} s^{-1}) in homoepitaxial CVD Ga_2O_3 was enabled by ultralow levels of background donor compensation by unintentional acceptors (2×10^{13} cm^{-3}, ~0.06% donor compensation), likely originating from unintentional formation of point defects in the lattice (*7*).

However, achieving growth of very thick (>30 μm) epitaxial β-Ga_2O_3 at this level of purity is extremely challenging, and its development is required to be competitive with SiC for the ultrahigh-power switching applications space. The understanding of defects in Ga_2O_3 epitaxy will have to advance over the next several years before high-voltage Ga_2O_3 devices could become commercially viable. Point defects, such as vacancies and their related complexes (such as vacancy-interstitial defects) as well as extended defects in thick epitaxial layers, currently inhibit Ga_2O_3 device dimensions. In general, defect characterization in Ga_2O_3 promises to be a rich research field that will also enable any Ga_2O_3 power electronics commercial enterprise hoping to break the 20-kV barrier with a useful device size to do so.

For power electronics, the development of p-type (hole-carrier) materials is necessary because holes in Ga_2O_3 form localized polarons that lead to a self-trapping phenomenon that limits their conduction (*8*). The absence of p-type conductivity in Ga_2O_3 presents a challenge for high–electric field management regardless of device geometry, and any practical solution will require innovation in heterogeneous integration not faced for previously developed semiconductors.

Unlike p-type semiconductors, such as SiC, gallium nitride (GaN), or diamond, WBG p-type nickel oxide (NiO) can be sputtered at room temperature and is thus benign for integration with Ga_2O_3 devices. Recent studies, such as the 8-kV NiO/Ga_2O_3 p-n diode demonstration by Zhang *et al.* (*9*), have shown that the lack of p-type conductivity in Ga_2O_3 could potentially be managed by integrating heterojunctions with the purpose of field management and charge balance in these devices. The prospects of Ga_2O_3 as a power electronics material would be greatly enhanced if robust heterogeneous integration with p-type WBG semiconductors (such as GaN or AlN) are developed. Such development could lead to a reliable junction barrier Schottky rectifier being commercialized, as was the case for SiC.

Effective electric field termination at surfaces is a key requirement for the use of UWBG materials in practical high-voltage electronic devices. The nitrogen deep acceptor has been effective in rendering Ga_2O_3 nearly insulating and creating an effective dielectric layer that can reduce electric fields. Selective ion implantation could be useful in device fabrication processes to pattern conducting and insulating surface regions. Dry etching—a common processing step for creating such patterns, which can introduce surface defects that affect device reliability—could potentially be eliminated altogether if patterning could be implemented solely through ion implantation. Unlike any other UWBG material, Ga_2O_3 can even be wet etched in phosphoric acid and etched using gas-phase Ga, both of which eliminate chemical and mechanical damage from plasma etching that always introduces defects at the etched surface (*10*, *11*). Developing Ga_2O_3-specific fabrication processes in parallel with high-quality thick epilayers could speed up Ga_2O_3 device commercialization during the next decade, at least at the scale of a two-terminal device (such as a diode) (*5*).

The extremely low thermal conductivity of Ga_2O_3 (11 to 27 W m^{-1} K^{-1}) must be carefully considered (*5*). Cooling of Ga_2O_3 transistors will be even more critical than that of GaN transistors, which also suffer from self-heating effects (*12*). Although Ga_2O_3 devices still output about an order of magnitude less power during operation compared with their GaN counterparts, the top- and substrate-side cooling approaches developed for GaN could be applied to Ga_2O_3 (*12*). Indeed, capping a lateral transistor with AlN or nanocrystalline diamond has allowed for the achievement of 5– to 6–W mm^{-1} DC output power with Ga_2O_3, which is similar to early results from GaN high–electron mobility transistors in the 1990s. Heterogeneous integration with high–thermal conductivity, WBG p-type semiconductors, such as SiC, GaN, and even diamond, is of particular interest for p-n and junction barrier Schottky rectifiers.

Looking to early commercialization efforts of WBG semiconductors, the success of SiC was driven in part by substantial government investment and scientific research efforts that continue its innovation. Solving the issues of micropipe and basal plane dislocation defects in SiC relied on advanced characterization techniques, such as ultraviolet photoluminescence imaging and spectroscopy. Material scientists have continued to develop their understanding of defects in the ever-larger-diameter SiC wafers (*13*).

Similar efforts will be needed to understand and control both point and extended defects in thick (>30 μm) Ga_2O_3 epitaxial layers (*14*). Government funding would be crucial in the early support of such efforts. The small-business technology transfer program initiated by the US Office of Naval Research in 2017 with the purpose of initiating the development of β-Ga_2O_3 CVD has resulted in the commercialization of this capability before the program's end and highlights the importance of this new technology. The recently enacted US CHIPS and Science Act will not only fund chip fabrication facilities but will also direct $13 billion to the US Department of Commerce and the US Department of Defense toward semiconductor and microelectronics research and development. These investments should spur additional funding of research into UWBG semiconductors and related materials in the coming years, with the expectation that a diverse portfolio of heterogeneously integrated semiconductor modules will overcome the drawbacks of chips made using one particular semiconductor. Furthermore, a higher-frequency device will be beneficial at the system level only if the passive components can keep up. Advances in magnetic materials could also help prevent components such as inductors and transformers from becoming too lossy at higher frequencies.

Exploratory research into Ga_2O_3 as well as other UWBG semiconductors will likely continue given that fundamental questions are still being resolved (*15*). For Ga_2O_3, the largest hurdles, such as scaling epilayer thickness without sacrificing material quality and managing ultralow thermal conductivity and unipolar conductivity, will require new physics to be solved as well as the development of new or modified materials. The continued pace of innovative engineering of Ga_2O_3 could bring commercialization of GaO_3 power diodes within the next decade, but only if substantial investments in this interesting semiconductor continue. ■

REFERENCES AND NOTES

1. C. R. Eddy Jr., D. K. Gaskill, *Science* **324**, 1398 (2009).
2. M. Higashiwaki, K. Sasaki, A. Kuramata, T. Masui, S. Yamakoshi, *Appl. Phys. Lett*. **100**, 013504 (2012).
3. S. J. Pearton *et al*., *Appl. Phys. Rev*. **5**, 011301 (2018).
4. Novel Crystal Technology, Inc., Current commercial -Ga2O3 wafer suppliers; www.novelcrystal.co.jp/.
5. A. J. Green *et al*., *APL Mater*. **10**, 029201 (2022).
6. J. A. Spencer *et al*., *Appl. Phys. Rev*. **9**, 011315 (2022).
7. G. Seryogin *et al*., *Appl. Phys. Lett*. **117**, 262101 (2020).
8. J. B. Varley, A. Janotti, C. Franchini, C. G. Van de Walle, *Phys. Rev. B* **85**, 081109 (2012).
9. J. Zhang *et al*., *Nat. Commun*. **13**, 3900 (2022).
10. H.-C. Huang, Z. Ren, C. Chan, X. Li, *J. Mater. Res*. **36**, 4756 (2021).
11. N. K. Kalarickal *et al*., *Appl. Phys. Lett*. **119**, 123503 (2021).
12. M. J. Tadjer, T. J. Anderson, Eds., *Thermal Management of Gallium Nitride Electronics* (Elsevier, 2022).
13. N. A. Mahadik, R. E. Stahlbush, W. Sung, *J. Appl. Phys*. **131**, 225702 (2022).
14. M. D. McCluskey, *J. Appl. Phys*. **127**, 101101 (2020).
15. J. Y. Tsao *et al*., *Adv. Electron. Mater*. **4**, 1600501 (2018).

ACKNOWLEDGMENTS

The views expressed in this article do not necessarily represent the views of the US Department of Defense or the United States. Research at the Naval Research Laboratory was supported by the Office of Naval Research. K. Hobart, C. Eddy Jr., R. Stahlbush, A. Jacobs, H. Masten, J. S. Lundh, and T. Anderson; epitaxial growth by Novel Crystal Technology (Sayama, Japan); and Agnitron Technology (Chanhassen, MN) are gratefully acknowledged.

10.1126/science.add2713

REVIEW

Carbon nanotube transistors: Making electronics from molecules

Aaron D. Franklin[1,2]*, Mark C. Hersam[3,4,5], H.-S. Philip Wong[6,7]

Semiconducting carbon nanotubes are robust molecules with nanometer-scale diameters that can be used in field-effect transistors, from larger thin-film implementation to devices that work in conjunction with silicon electronics, and can potentially be used as a platform for high-performance digital electronics as well as radio-frequency and sensing applications. Recent progress in the materials, devices, and technologies related to carbon nanotube transistors is briefly reviewed. Emphasis is placed on the most broadly impactful advancements that have evolved from single-nanotube devices to implementations with aligned nanotubes and even nanotube thin films. There are obstacles that remain to be addressed, including material synthesis and processing control, device structure design and transport considerations, and further integration demonstrations with improved reproducibility and reliability; however, the integration of more than 10,000 devices in single functional chips has already been realized.

Transistors are electronic switching devices that enable digital computation based on their on-state (binary 1) and off-state (binary 0) operation. In the earliest days of integrated circuits, it became clear that scaling down the size of the transistors would drive better chip-level performance, which is now known as Moore's law. One of the most important dimensions for such scaling is the semiconducting channel length, which is the distance that electrical current flows or is controlled by a gate electric field to turn the device on and off. Although the initial channel lengths were many microns in size, proposals to scale the semiconducting channel to the ultimate limit of molecular dimensions (fractions of a nanometer) date to the mid-1970s (*1*). Decades of study on the transfer of electrons through conjugated organic molecules, considered to replace the silicon channel, highlighted several important challenges for such molecular transistors. The foremost issues included low stability and the difficulty of effectively gating, and making reliable electrical contact to, the molecules (*2*).

To meet or exceed the performance of silicon electronics, it became clear that new channel materials must have similar stability. Among the molecular options, semiconducting single-walled carbon nanotubes (CNTs) have several advantages (*3*, *4*). Nested multiwalled CNTs are effectively metallic at room temperature and thus have limited utility as transistor channels (*5*). Throughout this review, CNTs will imply single-walled nanotubes. Semiconducting CNTs are composed of a cylindrical shell of hexagonally arranged carbon with a diameter of ~1 nm. Electrons travel only forward or backward with a wave function wrapping around the nanotube to create a one-dimensional (1D) semiconductor with an energy bandgap of a few hundred milli–electron volts (*5*). These materials are stable in air and can be manipulated through a variety of processing methods that are commonly used in the semiconductor industry. The early demonstrations of field-effect transistors (FETs) by draping a semiconducting CNT over metal electrodes (*3*, *4*) have led to continued research activity with the goal of creating reproducible, scalable, and high-performance devices integrated into dense circuitry using processing steps similar to those used to create silicon electronics.

The widespread interest in semiconducting CNTs has also inspired intense and ongoing exploration of other nanomaterials, including semiconducting nanowires (*6*), 2D graphene (*7*), transition metal dichalcogenides (*8*), and Xenes (*9*). Despite the growing number of nanomaterial options, CNTs stand out in offering stability, bandgap, and superb electrical and thermal properties that are unmatched by other candidates. Here, we review recent material, device, and technology advances for CNT transistors, establishing both the substantial promise and the remaining challenges for this molecular transistor. Progress in the field will be related to the foremost potential applications for CNT transistors, highlighted in Fig. 1. Two of the most prominent potential applications are high-performance (HP) computing chips and thin-film transistors (TFTs) for display backplanes and the internet of things (IoT); a few of the target performance metrics for these applications are summarized in Table 1.

[1]Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA. [2]Department of Chemistry, Duke University, Durham, NC, USA. [3]Department of Materials Science and Engineering, Northwestern University, Evanston, IL, USA. [4]Department of Chemistry, Northwestern University, Evanston, IL, USA. [5]Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL, USA. [6]Department of Electrical Engineering, Stanford University, Stanford, CA, USA. [7]Stanford SystemX Alliance, Stanford University, Stanford, CA, USA.
***Corresponding author. Email: aaron.franklin@duke.edu**

Advances in materials for CNT transistors

Exploiting the advantages of semiconducting CNTs requires overcoming several materials science hurdles. Just as silicon must be purified and doped to be a useful channel material, as-synthesized CNTs can be either metallic or semiconducting and must be purified into semiconducting-only for use in transistors. Whether CNTs are metallic or semiconducting depends on how the hexagonal lattice wraps into a tube. This structure is most easily visualized by rolling a rectangular section of the sp^2-bonded hexagonal carbon lattice of atomically thin graphene into a 1D cylinder, where the resulting diameter is ~1 nm and the length is 10^2 to 10^8 nm. The vector that defines the width of the rectangular section with respect to the graphene lattice is commonly referred to as the chiral vector and ultimately determines the diameter, helicity, and conductivity properties of the CNT (*5*).

In addition to specifying the physical structure of the CNT, the chiral vector also imposes well-defined quantum-mechanical boundary conditions on the electronic band structure, which implies that for random tube closure, ~33% of CNT chiralities are metallic and ~67% are semiconducting. Moreover, among the semiconducting chiralities, the bandgap is approximately inversely proportional to the CNT diameter. Because CNT transistors require semiconducting channels, preferably with a well-defined and uniform bandgap, the ability to scalably synthesize and isolate CNTs with atomically precise chiral vector control is the ultimate goal for high-performance CNT integrated circuits.

Controlled synthesis of CNTs

CNTs can be synthesized by introducing a carbonaceous feedstock with a metal catalyst (usually Fe or Ni) into a growth chamber, where energy is added through heat, light, or plasma excitation. Because CNT growth typically occurs at temperatures at which these catalysts undergo substantial restructuring, it is difficult to control the chiral vector, and a range of CNT diameters and both electronic types are produced; much effort has been expended to gain control over CNT chirality (*10*). These approaches include the use of refractory catalyst particles such as W-Co alloys with well-defined size and shape that remain structurally invariant at the growth temperature and thus can drive predictable nucleation of targeted CNT chiralities (Fig. 2A) (*11*), the addition of molecular seeds that have a structure that closely matches the targeted CNT chirality (*12*), or the deployment of CNTs themselves as seeds in "CNT cloning" (*13*). Although tailored catalysts or seeds help control synthetic outcomes, many other growth parameters also play a role—including temperature, pressure, flow rates, and applied electric fields (*14*)—and

thus growth optimization entails the search of a broad parameter space. In an effort to accelerate this exploration, autonomous growth using closed-loop iterative experimentation is showing promise for rapid identification of synthesis conditions that minimize CNT structural polydispersity (*15*).

Separation of semiconducting CNTs

Because the most-optimized CNT growth procedures still lack sufficient monodispersity for wafer-scale transistor applications, post-synthetic separation methods are required to sort as-grown CNTs by diameter, chirality, and electronic type. Fortunately, CNTs have sizes and shapes that are comparable to those of biological macromolecules, which has allowed many CNT separation methods to be adapted from ones that have already been developed for biochemistry. In density gradient ultracentrifugation (DGU), CNTs are first dispersed and encapsulated with mixtures of surfactants that show selectivity for different CNT separation targets (including chiral vector, chiral handedness, electronic type, and diameter) and then separated by buoyant density in aqueous density gradients (*16*). Although DGU has sufficient scalability to be viable commercially, other strategies from biochemistry have also been heavily developed, including gel chromatography (*17*) and dielectrophoresis. The latter method has the added benefit of enabling aligned assembly of CNTs between prepatterned electrodes (*18*).

Methods from polymer chemistry have also been used for CNT separations, including aqueous two-phase extraction (*19*) and selective dispersion of targeted CNT chiralities with structure-discriminating polymers that wrap the nanotubes (Fig. 2B) (*20*). In all cases, purities of semiconducting CNTs have reached the detectable limits of optical spectroscopic characterization (~99.9%) and begun to provide sufficient monodispersity for many CNT transistor applications. The ultimate goal for high-performance digital transistors is to achieve >99.9999% pure semiconducting CNTs (see Table 1)—the higher the purity, the better the corresponding performance. In addition, any molecular wrapper (e.g., surfactant or polymer) should ideally be completely removed after deposition of the CNTs because this presents an unwanted residue that can hamper electrical contact, gating efficiency, and transport in the CNT transistor.

Other material considerations

A transistor also requires electrical contacts, doping, and dielectrics. Because contacts from commonly used metals (e.g., Au, Pd) tend to yield Fermi-level alignment near the valence band of CNTs, p-type behavior from the injection of holes is readily achieved for CNT

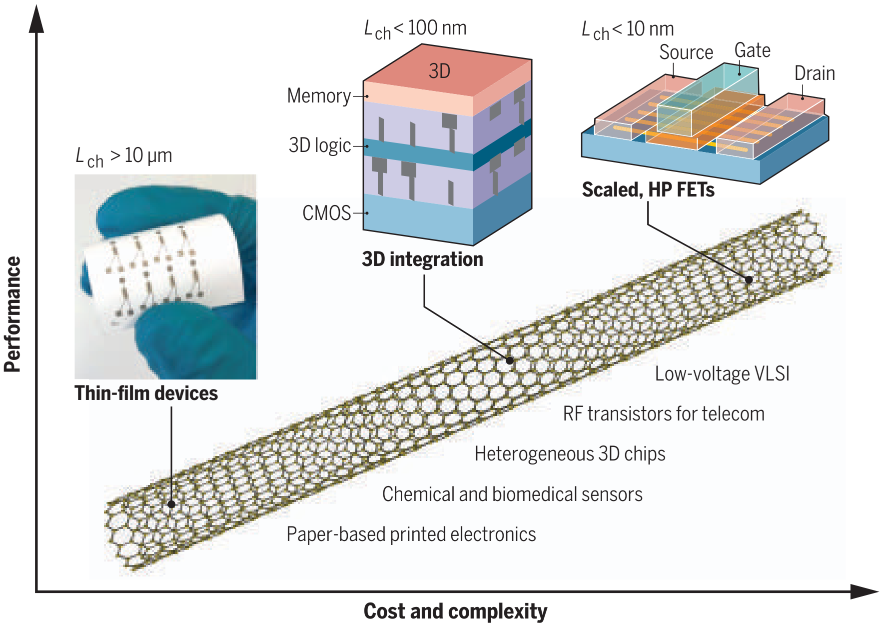


Fig. 1. Broad range of potential applications for CNT transistors. Illustration of device performance versus cost and complexity for some of the foremost potential applications of CNT transistors. Applications range from microscale thin-film devices (e.g., printed electronics, biosensors) to three-dimensionally integrated BEOL devices (such as heterogeneous 3D layers integrated onto silicon CMOS) and scaled high-performance (HP) FETs [such as low-voltage very-large-scale integration (VLSI)], with increasing performance corresponding with increased cost and complexity of integration. L_{ch}, channel length.

Table 1. A few of the target metrics for two prominent CNT transistor applications. Values are approximations based on achieving optimal performance. Notably, although some of these targets have been achieved, one of the foremost challenges is to achieve them simultaneously (e.g., high on-state current with low subthreshold swing, which is a measure of how much gate voltage is required to modulate the current by one decade). High-performance FETs are used in applications such as central processing units (CPUs) for servers, and TFTs are thin-film transistors such as those used in the backplane electronics of displays.

Metric	Target for high-performance FETs	Target for TFTs
CNT semiconducting purity	>99.9999%	>99.9%
CNT array alignment	Parallel CNTs, consistent pitch	Thin-film CNTs in an unaligned network
CNT array density	>200 CNTs per micrometer (linear density)	>50 CNTs per square micrometer (aerial density)
Channel length	<12 nm	>10 μm
Contact length	<10 nm	>1 μm
On-state current	> 0.5 mA μm^{-1} at 0.6 V	>100 μA mm^{-1} at 3 V
Contact resistance	<50 ohm•μm per side	<20 kilohm•μm per side
Subthreshold swing	<70 mV per decade	<200 mV per decade (application dependent)

transistors (*21*). However, the requirement for complementary p-type and n-type transistors in digital circuits implies that controlled n-type injection and/or doping is required. Electron-donating adsorbates, such as organorhodium compounds (*22*), coupled with atomic layer–deposited encapsulation layers (*23*), enable the fabrication of highly stable n-type CNT transistors (Fig. 2C). Charge-selective contacts based on metal work function, such as Pd for p-type injection and Sc for n-type injection, also enable complementary CNT transistors (*24*, *25*). Beyond the metal selection, interfacial material considerations and overall contact structure also play a role (see Fig. 2D for an exemplary end-bonded contact structure using Mo). Extension regions of a metal-oxide-semiconductor FET (MOSFET), which are between the source

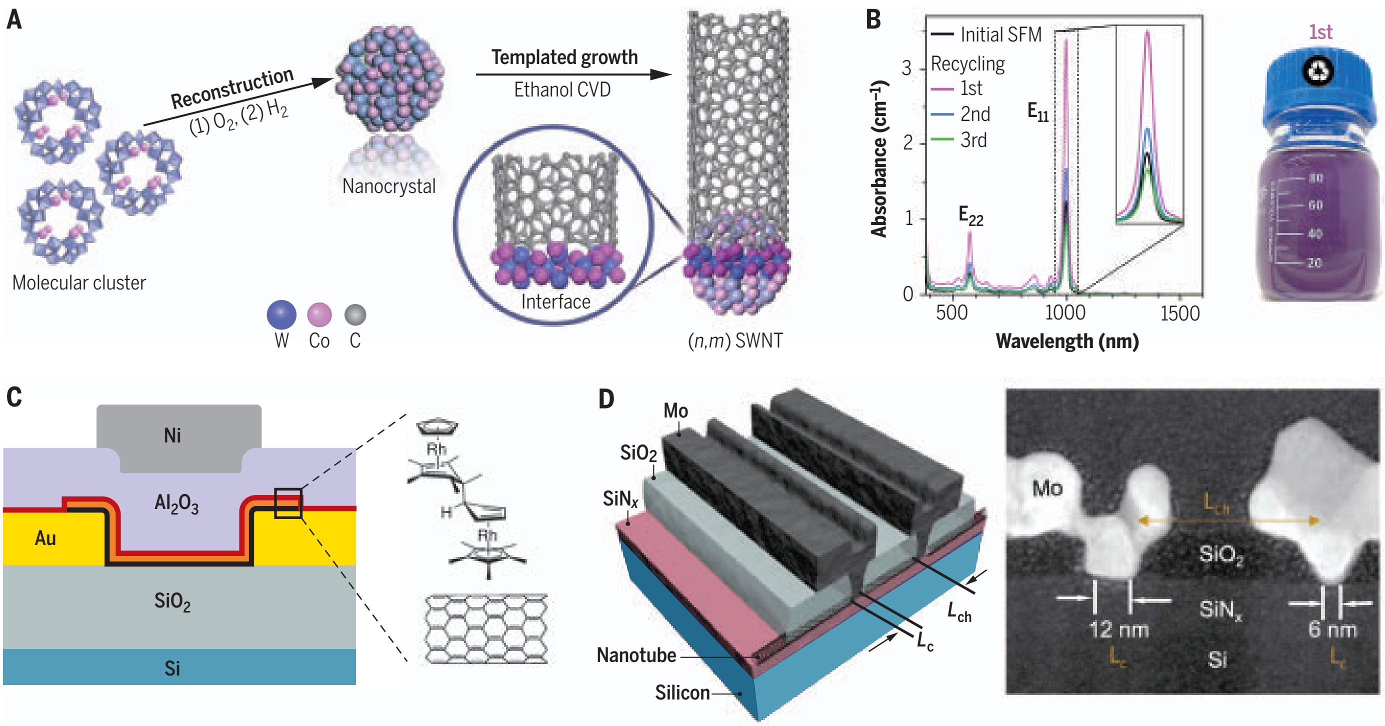


Fig. 2. Examples of materials for high-performance CNT transistors, including synthesized CNTs, purified CNT mixtures, doping strategies, and contact metals. (**A**) Templated CNT growth of targeted chiralities using refractory W-Co nanocrystal catalysts. CVD, chemical vapor deposition; SWNT, single-walled nanotube. [Adapted by permission from Springer Nature Customer Service Center GmbH, Springer *Nature* (*11*), copyright (2014)]. (**B**) Selective polymer dispersion enabling scalable isolation of targeted CNT chiralities from as-grown polydisperse mixtures as verified by absorption spectrum is shown on the left; a photo of the resultant bottle of sorted CNTs is shown on the right. E_{11} and E_{22}, absorption peaks; SFM, shear force mixing. [Adapted from (*20*) with permission from Elsevier]. (**C**) Electron-donating organorhodium compounds encapsulated with atomic layer–deposited alumina enabling stable n-type CNT transistors. Black is the CNT layer, orange is the dopant layer, and red is a seeding layer for dielectric growth. [Adapted with permission from (*22*). Copyright 2016 American Chemical Society]. (**D**) When reacted to form end-bonded carbides, molybdenum contacts to CNT transistors can be scaled down to sub-10-nm dimensions in contact length (L_c) while retaining efficient charge injection. [Adapted with permission from (*54*)].

or drain and the gated semiconducting channel, require stable doping with well-controlled doping levels that are optimized for the trade-off between series resistance and parasitic capacitances (*26*)—a feat yet to be reliably accomplished for CNT transistors. For the gate dielectric layer, specific materials such as Y_2O_3 have exhibited nearly ideal properties with a high dielectric constant κ and conformal dielectric coating on CNTs after oxidation of deposited yttrium (*27*). A more conventional approach that uses atomic layer deposition of Al_3O_3 and HfO_2 bilayer dielectrics has enabled transistors that have a 10-nm gate length with a gate leakage current commensurate with state-of-the-art Si transistors (*28*). Upon integration of all these optimized materials, CNT transistors have been shown to exceed the performance of incumbent silicon integrated circuit technology, as will be discussed in the subsequent sections.

CNT transistor design

The initial focus for CNT transistor research was on the use of a single CNT as the channel (see Fig. 3, A and B) and demonstration of ballistic transport (*21*) and digital circuit operability (*29*). Although devices with an individual nanotube channel are still of interest for sensing applications, they are no longer considered suitable for digital or radio-frequency (RF) electronics based on the need for higher current flow than a single CNT can deliver. Although the current-carrying capacity for CNTs is astonishing [~10^9 A cm^{-2} (*30*)], they are only ~1 nm in diameter, which yields only ~10 μA per CNT. Hence, recent work has predominantly focused on having multiple CNTs in the channel.

Aligned arrays of CNTs

Ideally, the CNTs in a transistor channel would be perfectly aligned in a parallel array with a controlled pitch of ~2 to 5 nm (*31*), similar to how fins of silicon are arranged in modern transistor technologies (FinFETs). Realizing such arrays continues to be a challenge. If the CNTs are too close (or bundled), it can create cross-talk (electric field screening) and effective gating issues (*32*). If the CNTs are too far apart, current density (current per transistor width) will be insufficient. For digital systems with a high density of CNT transistors, variations in the pitch between CNTs also deleteriously affects the overall energy, delay, and noise margin (*33*).

Recent progress is encouraging, including a small-scale demonstration with a controlled CNT pitch of ~10 nm using DNA-directed assembly (*34*). There are also wafer-scale, high-throughput strategies that use various forms of solution-phase assembly (also referred to as dimension-limited self-alignment or liquid crystalline interfacial assembly), which achieved a ~20-nm pitch (Fig. 3, C and D) in one report (*35*) and a 5- to 10-nm pitch in another (*36*). The primary differences in the two studies were the polymer used to wrap the CNTs and the solution-phase technique of depositing the CNTs into arrays on the substrate. Nevertheless, these approaches still require further work to remove unwanted residue from the solution-phase processing along with more consistent, controlled alignment (without bundling) in all directions with uniform spacing.

Thin films of CNTs

The difficulty of achieving aligned arrays with controlled pitch has led some researchers to use unaligned CNT networks or thin films (Fig. 3, E to H). Although these unaligned films are less favorable for carrier transport, as well as for contacting and gating the nanotubes,

unaligned CNT networks have achieved high performance in nanoscale transistors (*37*, *38*). Moreover, CNT thin films can be deposited by using printing techniques, including roll-to-roll (*39*) and direct-write (*40*, *41*) approaches (Fig. 3H), which makes them attractive for TFTs. The application space for these larger (approximately tens of micrometers) TFTs is distinct from high-performance nanoscale FETs and includes sensors, flexible electronics, IoT, and display backplanes (*42*). For TFT applications, CNT thin films compete well against incumbent semiconductor options such as organics and polymers, metal oxides, and low-temperature polysilicon (LTPS) (*43*).

When CNT thin films are used in FETs with nanoscale channel lengths (<100 nm), most of the nanotubes bridge the entire channel, even if they are not perfectly aligned (Fig. 3F). In the microscale lengths of TFTs, nanotubes in the thin-film channel are not long enough to transverse the channel and instead operate as percolating networks in which electrons travel from CNT to CNT in transit from source to drain (Fig. 3G) (*44*, *45*). Compared with long-studied organic semiconductor TFTs (*46*), CNT-TFTs have considerably higher mobilities (10 to 100 $cm^2 V^{-1} s^{-1}$) and stability under bias, in air, or both.

Advanced gating structures

In addition to the density and arrangement of nanotubes in the channel, the gate configuration in a CNT transistor has advanced in many ways. For nanoscale FETs, the primary goal is to maximize the gate control of the CNT energy bands in the channel, which is achieved through strong gate coupling that is typically expressed as a small scale length, λ (*47*). The scale length depends on the gate geometry and on the thickness and permittivity of both the gate dielectric and the semiconducting channel. A generally accepted approximation is that a channel length greater than 3λ will ensure that deleterious short-channel effects are avoided.

Given their intrinsically small size, CNTs offer advantages for aggressively scaled devices. Although it is ideal for an FET to have a gate-all-around geometry to minimize λ, and demonstrations of such gate structures for CNTs have been reported (*48*, *49*), studies have shown that channel lengths that are much less than 10 nm (as short as 5 nm) can be achieved in either bottom-gate (*50*, *51*) or top-gate (*51*, *52*) geometries. Although gate geometry does vary for TFTs, it is less critical and mostly limited by the gate dielectric material and the application needs.

Source-drain contact structures

For highly scaled CNT transistors with small footprints, not only does the channel length need to be at the nanometer scale but the source and drain contacts also need to have minimal dimensions while still providing efficient ohmic charge injection. Palladium contacts have achieved the quantum limit of 6.5 kilohm per CNT at a 10-nm contact length for a p-type side contact, where the metal rests on top of a CNT without any chemical bonding (*53*), though this needs to be realized with higher yield and reproducibility. Alternatively, an edge-contact structure would offer ideal scalability and has been demonstrated by reacting Mo with CNTs to yield a carbide end-bonded contact with sub-10-nm contact lengths (Fig. 2D) (*54*). Regardless of the geometry, contacts to CNTs are a leading factor in determining overall performance, and the combination of material, structure, and processing must be further refined to yield contacts for both p- and n-type carrier injection with high consistency and low resistance.

Technology demonstrations

High-performance, energy-efficient digital logic

Although many applications can benefit from the properties of CNTs, digital logic applications have received the greatest attention (Fig. 4) because they have the potential to surpass incumbent Si technology in performance and energy efficiency. Such exemplary high-performance devices from aligned arrays of CNTs can achieve high on-state currents at relatively low voltages (Fig. 4, A to C). As shown in Fig. 4D, gate-all-around CNT transistors with doped extensions and multiple layers of high-density CNTs are projected to show up to seven times the energy-delay product (EDP) benefits compared with Si nanosheets at the 2-nm technology node (the EDP, or switching energy, is the product of the time and the power consumption for an on-off cycle, and a measure of energy efficiency) (*26*). As noted

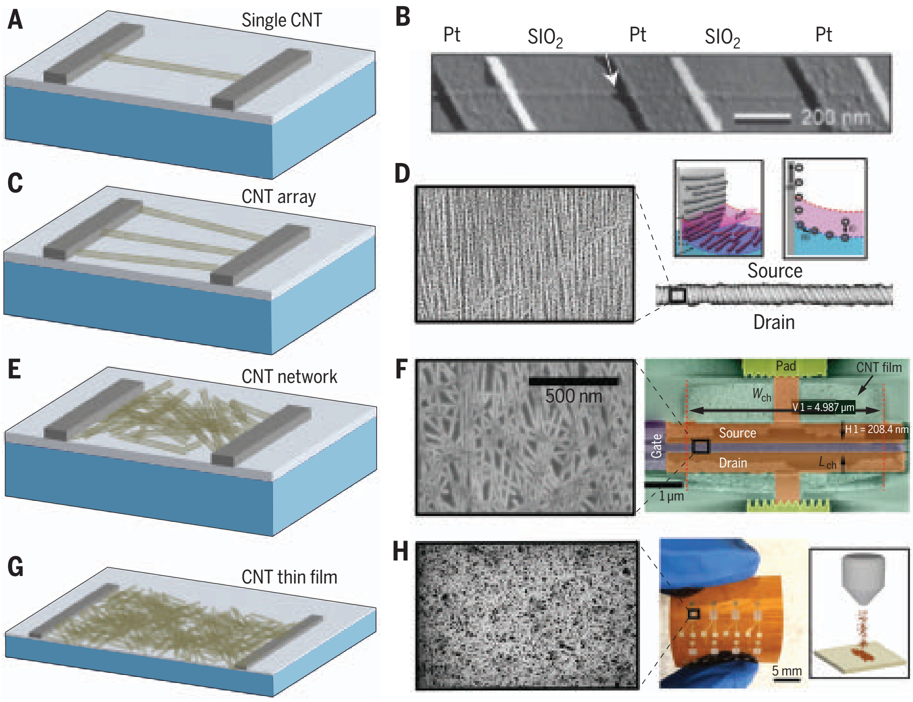


Fig. 3. Variations in CNT transistor structures. (**A**) Illustration of a single CNT channel with metallic source and drain contacts. (**B**) Atomic force microscopy image of the first-reported CNT transistor, which had a single nanotube. [Reprinted by permission from Springer Nature Customer Service Center GmbH, Springer *Nature* (*3*), copyright (1998)]. (**C** and **D**) Illustration of an aligned array of CNTs as the channel (C) and a corresponding recent example of a transistor with such an array (D), including a scanning electron microscopy (SEM) image of the aligned CNTs (left) and schematic of the solution-phase assembly process (right). [(D) is reprinted with permission from (*35*); Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC). https://creativecommons.org/licenses/by-nc/4.0/]. (**E** and **F**) Illustration of a CNT network (not aligned) used as channel for nanoscale FET (E) with a corresponding recent example of a high-performance transistor (F), including a SEM image of the high-density film (left) and a top-view schematic of the device structure (right). Note that most nanotubes directly bridge the source and drain in this nanoscale configuration. H1 and V1, measurement markers of CNT film area height and channel length, respectively; W_{ch}, width of CNT channel region. [(F) is adapted by permission from Springer Nature Customer Service Center GmbH, Springer *Nature* (*38*), copyright (2018)]. (**G** and **H**) Illustration of a CNT thin film used in a thin-film transistor (dimensions of tens of micrometers) (G) with a corresponding recent example of an aerosol-jet printed CNT-TFT on a flexible plastic substrate (H), including a SEM image of the printed thin film (left) and a picture of the printed CNT-TFTs with a schematic of the printing technique (right). [(H) is reprinted with permission from (*41*). Copyright 2019 American Chemical Society].

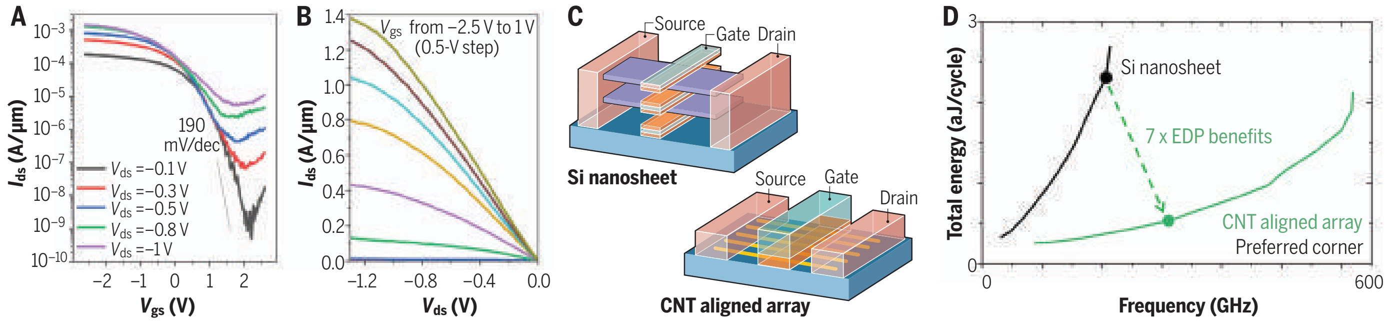


Fig. 4. High-performance CNT transistors for digital logic applications. (**A** and **B**) Subthreshold (A) and output characteristics (B) of CNT transistors fabricated with aligned arrays of ~150 CNTs per micrometer that achieve an on-state current of >1 mA μm^{-1}. I_{ds}, drain current; V_{ds}, drain-source voltage; V_{gs}, gate-source voltage. [Adapted with permission from (*36*)]. (**C**) Device schematics for a Si nanosheet transistor with two stacked channels and a CNT aligned array transistor. (**D**) Projected energy versus frequency pareto curves for Si nanosheet and CNT transistors at the 2-nm technology node for an inverter ring oscillator. [© 2021. Adapted, with permission, from (*26*)].

earlier, because of their ultrathin body (~1 nm), CNT transistors offer excellent electrostatic control even at aggressively scaled gate lengths, limited only by direct source-to-drain tunneling. Parasitic capacitance, a key detractor of speed and energy efficiency, accounts for >70% of the total capacitance of modern Si transistors. Because of the ultrathin body, CNT transistors have substantially lower parasitic gate-to-source or gate-to-drain capacitance. These two key attributes of CNTs, along with the high transport and injection velocities, are the physical basis for high-performance, energy-efficient digital logic.

As noted earlier, many fundamental building blocks of a CNT transistor technology have already been demonstrated. At the circuit or system level, a fully functional static random-access memory (SRAM) array (*55*), a monolithic 3D imager (*56*), and a 16-bit RISC-V (where RISC is reduced instruction set computer) processor with >14,000 transistors (Fig. 5B) (*57*) have been fabricated entirely from CNT transistors. What's more, wafer-scale fabrication of CNT transistors has been demonstrated in an industrial foundry using 200-mm wafer processing techniques (Fig. 5A) (*58*). The fabrication and design of CNT transistors with the same tools and infrastructure as commercial semiconductor technologies helps lower the barrier for the introduction of CNT devices into mass production.

At the individual device level, recent work shows short gate length (10 nm), complementary p- and n-channel devices with near ideal subthreshold swing for single-CNT transistors (*59*), and high on-state current per width for aligned CNTs with a density of 50 CNTs per micrometer (*60*). In the near future, it will be possible to integrate the following elements (already shown separately) in a single device demonstration: gate-all-around geometry (*48*, *49*), >250 CNTs per micrometer in highly aligned arrays (*36*), 3-nm oxide dielectric (target oxide capacitance = 2.94×10^{-10} F m^{-1}) (*28*), sub-10-nm p-type contacts with a contact resistance of 6.5 kilohm per CNT (*53*), sub-10-nm gate length (*52*), multiple stacked CNT channel layers (*61*), and doped source or drain extensions (*62*). This MOSFET-like CNT structure with 35-nm contacted gate pitch and 20-nm active width is projected to have a performance that far exceeds that of Si transistors for a 2-nm node logic technology.

3D integration

Future semiconductor chips will go beyond 2D device miniaturization and instead will have 3D layers of active devices (*63*). Because logic device layers in 3D must be thin and fabricated at temperatures that are compatible with back-end-of-line (BEOL) wiring layers (typically <400°C), CNT transistors are particularly well-suited for 3D integration because of the low device-fabrication temperature and thin device layer. Starting from the first demonstration of an all-CNT transistor computer almost a decade ago (*64*), progress has occurred not only in the level of integration but also in the variety of devices as well as maturation of the technology from university laboratories to industry.

A four-layer monolithically integrated chip comprising a silicon transistor layer, a CNT transistor memory read-out circuit layer, a resistive switching metal-oxide random-access memory (RRAM) layer, and a CNT transistor sensor layer on the top illustrates the benefits of monolithic integration (Fig. 5, C to E) (*65*). This 3D chip can process information from the sensors to the memory cells to the transistors in parallel at rates of terabytes per second. Another example is an end-to-end brain-inspired hyperdimensional computing nanosystem that is effective for cognitive tasks such as language recognition, which was realized with monolithic 3D integration of CNT transistors and RRAM, enabling fine-grained and dense vertical connections between computation and storage layers using BEOL interlayer vias (*66*). The CNT transistor fabrication process not only has been shown on full 200-mm wafers (*58*) but also has 3D integration with RRAM (*67*).

RF electronics

Although digital electronics remains the dominant focus in the field, CNT transistors also hold great promise for high-frequency RF transistors, which are relevant to telecommunications applications (*68*, *69*). Many of the material and device needs for digital CNT transistors also apply to RF electronics, with some relaxing of the semiconducting purity needs and an enhanced need for high transconductance and linearity, which translates to low distortion when amplifying a signal. Recent progress on RF CNT transistors from aligned arrays of nanotubes shows the ability to operate at frequencies up to hundreds of gigahertz with attractively low power consumption and high versatility for integration in system-on-chip applications (*70*).

Printed electronics

The ability to purify a solution-phase dispersion of semiconducting CNTs also enables printing into thin-film devices (Fig. 2H). Many reports have shown that fully printed CNT-TFTs can be used in digital logic circuitry to illustrate the ability of these devices to deliver computational functionality (*71*–*73*). However, given the low cost of legacy-node silicon transistor technologies, the likelihood that printed CNT-TFT circuitry will be of widespread use is low. More encouraging is the use of printed CNT-TFTs for the backplane control of displays (*74*) or for custom biosensing systems (*75*). Recent studies also reveal the recyclability of CNT thin films (*76*), which shows promise for enabling a fully printed, paper-based electronic system in which all core materials are able to be recaptured and reused (*77*).

Future developments and perspectives

Materials outlook

Advances in materials are anticipated to be central to future advances in CNT transistors. Improving the purity of semiconducting CNTs is critical for all device use cases. In this regard, one of the largest impediments to minimizing

metallic CNT impurities down to concentrations of parts per million or billion is the lack of high-throughput analytical methods for detecting ultralow concentrations of metallic CNTs. Most high-throughput optical detection methods for CNTs (such as photoluminescence spectroscopy) are less sensitive, if not completely insensitive, to metallic species. Indeed, the only established method for quantifying ultralow concentrations of metallic CNTs is to fabricate massive arrays of individual CNT transistors and then electrically probe them one by one in search of short circuits. This approach is extremely time consuming and only gets worse as the semiconducting purity increases. Thus, most CNT separation methods have only been optimized to the detection limits of optical spectroscopy (~99.9%).

Another unresolved issue for semiconducting CNTs is the need for a scalable and sustainable manufacturing approach to produce sufficient quantities of ultrahigh-purity semiconducting CNTs to meet the potentially large market represented not only by high-performance integrated circuits but also by high-volume printed electronics. Most solution-based separation methods do not possess fundamental barriers to scalability, but the yields of these processes are ultimately limited by the quality of the input raw material. Improvements in synthesis that minimize impurities and maximize semiconducting purity with narrow CNT diameter distributions are needed to improve the yield of downstream separation. An enticing option would be to refine cloning (*13*) to the point that iterative separation and amplification could be achieved in a manner analogous to the polymerase chain reaction (PCR) in biochemistry.

Ultimately, growth conditions encompass such a vast parameter space that methods for efficiently searching for and identifying optimal growth conditions are needed. Emerging artificial intelligence and machine-learning optimization approaches coupled with high-throughput experimental screening hold promise for next-generation synthetic efforts (*15*). Similarly, the discovery, optimization, and integration of the many other materials in a CNT transistor (including dopants, contacts, gate electrodes, and dielectrics) can also likely be accelerated by machine learning coupled with high-throughput experimental screening.

Device outlook

Although much has been learned about establishing interfaces to CNTs, including gate structure and contacts, challenges remain. The roles of material selection and purification (discussed earlier), methods of fabrication, and doping control continue to be elucidated in an expansive volume of reports. Indeed, one of the foremost challenges moving forward is determining what combination of materials and processes (of the thousands reported) is most appropriate to use. More systematic studies are needed that explore certain contact and gate-stack material configurations for their impact on device performance, yield, reproducibility, and stability. For example, it is clear that CNT channels are scalable to sub-10-nm lengths in a variety of configurations, but it is not clear which device structure is superior (e.g., top-gate versus gate-all-around, side contact versus edge contact) and whether top-performing options also have fabrication processes that are compatible with relevant manufacturing in complementary metal-oxide-semiconductor (CMOS) fabs. Most metal-contact formation processes rely on liftoff, which is not considered a scalable process, and the liftoff-free alternatives also tend to rely on slow patterning processes (*78*).

The scalability of the contact length, which is an equally important parameter as the gate length for overall transistor scaling, needs further consideration. Some studies show severe degradation at sub-30-nm contact lengths (*52*), whereas others have shown less degradation at scaled lengths but have not yet realized them at high yield (*53*). Such contact-length scaling challenges are common to all transistors (*79*), but discovering a solution that allows for aggressively scaled contacts without degrading the device would be a critical advance. End-bonded or edge contacts present one such possibility (*54*), although further work is required to reduce the processing temperature and to understand transport and performance limits. In addition, realizing an equally high-quality and scalable contact to n-type CNT transistors remains to be addressed.

Regarding TFTs from CNTs, much of the knowledge gained from nanoscale FET devices is applicable. The foremost exceptions are that a TFT technology should ideally be compatible with large substrate sizes and have exceptionally low cost. Because one of the primary applications for TFTs is in display backplanes, the materials and processes should be scalable to large panels. Although device-level performance and size matter, TFTs have relaxed constraints, with more emphasis given to fabrication cost because these devices will be used in commodity applications (such as backplanes) or disposable applications (such as IoT). The recent demonstration of recyclable printed

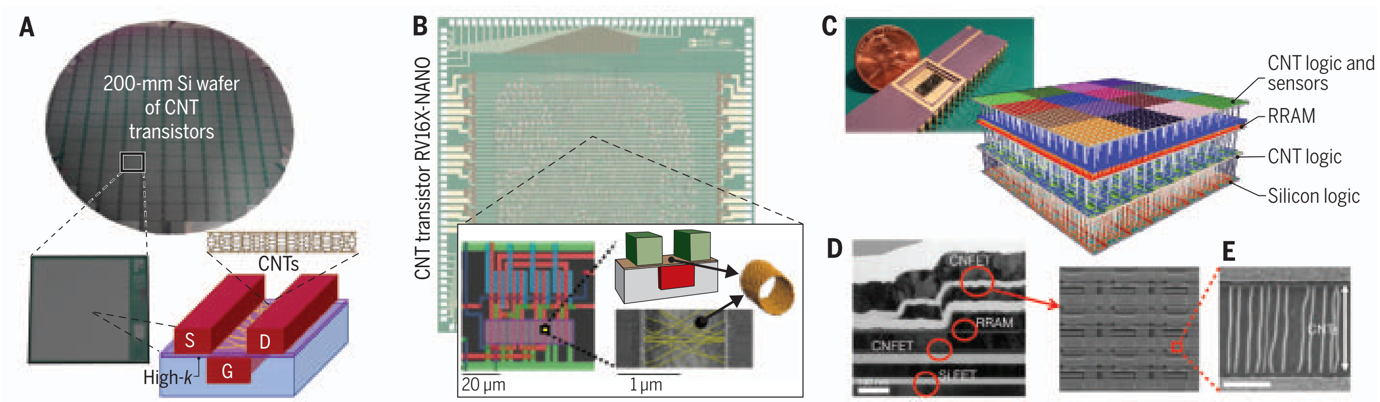


Fig. 5. Wafer-scale and 3D integration of CNT transistors. (**A**) A 200-mm Si wafer with CNT transistors processed in a commercial silicon foundry. An image of a single die or chip from the wafer is shown at the bottom left, and a schematic of the CNT transistor structure is shown at the bottom right. D, drain; G, gate; *k*, relative permittivity; S, source. [Adapted by permission from Springer Nature Customer Service Center GmbH, Springer *Nature* (*58*), copyright (2020)]. (**B**) Optical image of a RISC-V processor realized with CMOS CNT transistors (RV16X-NANO), including higher magnification images showing details of CNT circuits (false colors represent different metal layers) and a single CNT device (CNTs are highlighted in yellow). [Adapted by permission from Springer Nature Customer Service Center GmbH, Springer *Nature* (*57*), copyright (2019)]. (**C** to **E**) Image and schematic of a 3D N3XT chip with monolithic integration of CNT transistors and RRAM memory layers on top of silicon logic (C); cross-sectional TEM image showing the bottom Si logic layer, the RRAM memory layer, and the two CNT transistor layers [carbon nanotube field-effect transistor (CNFET), logic, and sensors] (D); and scanning electron microscopy images of a CNT circuit and devices in the top layer of the 3D N3XT chip (E) (scale bar, 500 nm). [Adapted by permission by Springer Nature Customer Service Center GmbH, Springer *Nature* (*65*), copyright (2017)].

CNT-TFTs on paper substrates suggests sustainable implementations (*77*). Improvements in CNT-TFT yield and stability will be critical, particularly the role of tube-tube contacts in percolating networks.

Technology outlook

The realization of a CNT transistor technology that meets high-volume manufacturing needs has many remaining hurdles that require a concerted effort from academia and industry to surmount. Regarding semiconducting CNT purity, although the highest possible purity remains ideal for EDP, logic design techniques can be used to relax the requirement for certain applications by about 100 times (from 99.9999 to 99.99%), without imposing additional processing steps or redundancy (*65*). For high-performance digital systems, device variations play an important role in determining the overall EDP and noise margin of the system. CNT-specific sources of variation include CNT density and pitch (distance between CNTs in a multi-CNT transistor), CNT bandgap (determined by the chirality and diameter), and extreme sensitivity to random fixed charges in the surroundings (which is also the reason why CNTs are ultrasensitive sensors). The transistor width (perpendicular to the direction of current flow) of logic technologies is on the order of 20 to 40 nm. For a CNT density of 250 CNTs per micrometer, there will only be 5 to 10 CNTs in the channel; hence, variations of the CNT density and the CNT pitch will lead to substantial variations in the current drive.

Design solutions that mitigate such variations are essential as part of a co-design process for technology development (*80*). For example, CNT bandgap variation directly translates into variation in off-state leakage current through the threshold voltage and band-to-band tunneling at the drain. Band-to-band tunneling leakage varies exponentially with bandgap and sets a minimum achievable leakage current (*81*)—a boundary for trading on-state current with off-state leakage current by tuning the threshold voltage. Direct source-to-drain tunneling current also depends exponentially on the bandgap and sets the limit for gate-length scaling. The choice of the CNT diameter (bandgap) is faced with the same trade-off as other FETs. Small-bandgap CNTs have lower effective masses and higher on-state currents, and large-bandgap CNTs have lower tunneling off-state leakage currents and can scale down further in gate length and maintain higher operating voltages for high speed. The optimal choice is necessarily application-dependent and must be co-designed, given the target computing workloads (*82*).

Although the CNT transistor inherits all the limitations of a MOSFET (electrostatics and transport physics) and has all the challenges of a low-dimensional channel material (contacts and surfaces without dangling bonds), it also retains all the benefits of an FET, which include a well-developed circuit or system design ecosystem and a mature manufacturing technology, with the further potential to integrate in three dimensions for chips with increasing device count and connectivity. These benefits are projected to eventually outweigh all the limitations because the power of incumbency and scalability in three dimensions cannot be understated. Between the opportunities in high-performance digital logic with the potential for 3D integration and the possibilities for printed and even recyclable thin-film electronics, CNT transistors warrant a renewed and even redoubled effort from academic, government, and industry contributors. These molecular transistor technologies are within reach, but only if the remaining challenges are surmounted by the scientific and engineering communities.

REFERENCES AND NOTES

1. A. Aviram, M. A. Ratner, *Chem. Phys. Lett.* **29**, 277–283 (1974).
2. M. Ratner, *Nat. Nanotechnol.* **8**, 378–381 (2013).
3. S. J. Tans, A. R. M. Verschueren, C. Dekker, *Nature* **393**, 49–52 (1998).
4. R. Martel, T. Schmidt, H. Shea, T. Hertel, P. Avouris, *Appl. Phys. Lett.* **73**, 2447–2449 (1998).
5. M. S. Dresselhaus, G. Dresselhaus, P. Avouris, *Carbon Nanotubes: Synthesis, Structure, Properties, and Applications* (Springer, 2001).
6. C. Jia, Z. Lin, Y. Huang, X. Duan, *Chem. Rev.* **119**, 9074–9135 (2019).
7. D. Akinwande *et al.*, *Nature* **573**, 507–518 (2019).
8. S. Das *et al.*, *Nat. Electron.* **4**, 786–799 (2021).
9. L. Zhang *et al.*, *Adv. Funct. Mater.* **31**, 2005471 (2021).
10. R. Rao *et al.*, *ACS Nano* **12**, 11756–11784 (2018).
11. F. Yang *et al.*, *Nature* **510**, 522–524 (2014).
12. J. R. Sanchez-Valencia *et al.*, *Nature* **512**, 61–64 (2014).
13. B. Liu *et al.*, *Nano Lett.* **13**, 4416–4421 (2013).
14. J. Wang *et al.*, *Nat. Catal.* **1**, 326–331 (2018).
15. P. Nikolaev *et al.*, *npj Comput. Mater.* **2**, 16031 (2016).
16. M. S. Arnold, A. A. Green, J. F. Hulvat, S. I. Stupp, M. C. Hersam, *Nat. Nanotechnol.* **1**, 60–65 (2006).
17. H. Liu, D. Nishide, T. Tanaka, H. Kataura, *Nat. Commun.* **2**, 309 (2011).
18. Q. Cao, S. J. Han, G. S. Tulevski, *Nat. Commun.* **5**, 5071 (2014).
19. J. A. Fagan *et al.*, *Adv. Mater.* **26**, 2800–2804 (2014).
20. A. Graf *et al.*, *Carbon* **105**, 593–599 (2016).
21. A. Javey, J. Guo, Q. Wang, M. Lundstrom, H. Dai, *Nature* **424**, 654–657 (2003).
22. M. L. Geier, K. Moudgil, S. Barlow, S. R. Marder, M. C. Hersam, *Nano Lett.* **16**, 4329–4334 (2016).
23. M. L. Geier *et al.*, *Nat. Nanotechnol.* **10**, 944–948 (2015).
24. Z. Chen, J. Appenzeller, J. Knoch, Y. Lin, P. Avouris, *Nano Lett.* **5**, 1497–1502 (2005).
25. C. Wang, K. Ryu, A. Badmaev, J. Zhang, C. Zhou, *ACS Nano* **5**, 1147–1153 (2011).
26. C. Gilardi *et al.*, in *2021 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2021), pp. 27.3.1–27.3.4.
27. Z. Wang *et al.*, *Nano Lett.* **10**, 2024–2030 (2010).
28. G. Pitner *et al.*, in *2020 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2020), pp. 3.5.1–3.5.4.
29. Z. Chen *et al.*, *Science* **311**, 1735 (2006).
30. B. Q. Wei, R. Vajtai, P. M. Ajayan, *Appl. Phys. Lett.* **79**, 1172–1174 (2001).
31. A. D. Franklin, *Nature* **498**, 443–444 (2013).
32. F. Léonard, *Nanotechnology* **17**, 2381–2385 (2006).
33. J. Zhang, N. P. Patil, A. Hazeghi, H.-S. P. Wong, S. Mitra, *IEEE Trans. Comput. Aided Des. Integrated Circ. Syst.* **30**, 1103–1113 (2011).
34. W. Sun *et al.*, *Science* **368**, 874–877 (2020).
35. K. R. Jinkins *et al.*, *Sci. Adv.* **7**, eabh0640 (2021).
36. L. Liu *et al.*, *Science* **368**, 850–856 (2020).
37. C. Zhao *et al.*, *Adv. Funct. Mater.* **29**, 1808574 (2019).
38. D. Zhong *et al.*, *Nat. Electron.* **1**, 40–45 (2018).
39. P. H. Lau *et al.*, *Nano Lett.* **13**, 3864–3869 (2013).
40. P. Chen *et al.*, *Nano Lett.* **11**, 5301–5308 (2011).
41. S. Lu *et al.*, *ACS Nano* **13**, 11263–11272 (2019).
42. A. D. Franklin, *Science* **349**, aab2750 (2015).
43. J. A. Cardenas, J. B. Andrews, S. G. Noyce, A. D. Franklin, *Nano Futures* **4**, 012001 (2020).
44. N. F. Zorn, J. Zaumseil, *Appl. Phys. Rev.* **8**, 041318 (2021).
45. Q. Cao *et al.*, *Nature* **454**, 495–500 (2008).
46. S. Park, S. H. Kim, H. H. Choi, B. Kang, K. Cho, *Adv. Funct. Mater.* **30**, 1904590 (2020).
47. R. Yan, A. Ourmazd, K. F. Lee, *IEEE Trans. Electron Dev.* **39**, 704–710 (1992).
48. Z. H. Chen *et al.*, *IEEE Electron Device Lett.* **29**, 183–185 (2008).
49. A. D. Franklin *et al.*, *Nano Lett.* **13**, 2490–2495 (2013).
50. A. D. Franklin *et al.*, *Nano Lett.* **12**, 758–762 (2012).
51. Q. Cao, J. Tersoff, D. B. Farmer, Y. Zhu, S. J. Han, *Science* **356**, 1369–1372 (2017).
52. C. Qiu *et al.*, *Science* **355**, 271–276 (2017).
53. G. Pitner *et al.*, *Nano Lett.* **19**, 1083–1089 (2019).
54. Q. Cao *et al.*, *Science* **350**, 68–72 (2015).
55. P. S. Kanhaiya, C. Lau, G. Hills, M. D. Bishop, M. M. Shulaker, *IEEE Trans. Electron Dev.* **66**, 5375–5380 (2019).
56. T. Srimani, G. Hills, C. Lau, M. Shulaker, in *2019 Symposium on VLSI Technology* (TEL, 2019), pp. T24–T25.
57. G. Hills *et al.*, *Nature* **572**, 595–602 (2019).
58. M. D. Bishop *et al.*, *Nat. Electron.* **3**, 492–501 (2020).
59. L. Peng, Z. Zhang, C. Qiu, *Nat. Electron.* **2**, 499–505 (2019).
60. G. J. Brady *et al.*, *Sci. Adv.* **2**, e1601240 (2016).
61. Y. Xie, Z. Zhang, D. Zhong, L. Peng, *Nano Res.* **12**, 1810–1816 (2019).
62. C. Lau, T. Srimani, M. D. Bishop, G. Hills, M. M. Shulaker, *ACS Nano* **12**, 10924–10931 (2018).
63. S. Salahuddin, K. Ni, S. Datta, *Nat. Electron.* **1**, 442–450 (2018).
64. M. M. Shulaker *et al.*, *Nature* **501**, 526–530 (2013).
65. M. M. Shulaker *et al.*, *Nature* **547**, 74–78 (2017).
66. T. F. Wu *et al.*, in *2018 IEEE International Solid - State Circuits Conference (ISSCC)* (IEEE, 2018), pp. 492–494.
67. T. Srimani *et al.*, in *2020 IEEE Symposium on VLSI Technology* (IEEE, 2020), pp. 1–2.
68. Y. Cao *et al.*, *ACS Nano* **10**, 6782–6790 (2016).
69. M. Schroter, M. Claus, P. Sakalas, M. Haferlach, W. Dawei, *Electron Devices Soc. IEEE J.* **1**, 9–20 (2013).
70. H. Shi *et al.*, *Nat. Electron.* **4**, 405–415 (2021).
71. M. Ha *et al.*, *ACS Nano* **4**, 4388–4395 (2010).
72. S. Lu, A. D. Franklin, *Nanoscale* **12**, 23371–23390 (2020).
73. K. Hong *et al.*, *Adv. Mater.* **26**, 7032–7037 (2014).
74. K. Schnittker, M. Tursunniyaz, J. B. Andrews, *J. Inf. Disp.* **22**, 193–209 (2021).
75. Z. Lin, G. Wu, L. Zhao, K. W. C. Lai, *IEEE Nanotechnol. Mag.* **13**, 4–14 (2019).
76. S. Wang, J. Zhao, Q. Wang, D. Zhang, *ACS Sustain. Chem. Eng.* **10**, 3851–3861 (2022).
77. N. X. Williams, G. Bullard, N. Brooke, M. J. Therien, A. D. Franklin, *Nat. Electron.* **4**, 261–268 (2021).
78. A. J. M. Mackus *et al.*, *Appl. Phys. Lett.* **110**, 013101 (2017).
79. S. K. Su *et al.*, in *2022 IEEE Symposium On VLSI Technology and Circuits* (IEEE, 2022), pp. 403–404.
80. G. Hills *et al.*, *IEEE Trans. Comput. Aided Des. Integrated Circ. Syst.* **34**, 1082–1095 (2015).
81. Q. Lin *et al.*, *IEEE Electron Device Lett.* **43**, 490–493 (2022).
82. G. Hills *et al.*, *IEEE Trans. NanoTechnol.* **17**, 1259–1269 (2018).

ACKNOWLEDGMENTS

Funding: A.D.F. acknowledges support from the National Science Foundation (NSF) (NSF ECCS-1915814). M.C.H. acknowledges support from the NSF Materials Research Science and Engineering Center at Northwestern University (NSF DMR-1720139). H.-S.P.W. acknowledges support in part from the Defense Advanced Research Projects Agency (DARPA) Electronics Resurgence Initiative (ERI) Three Dimensional Monolithic System on a Chip (3DSoC), the NSF (award 1640060), and the Stanford SystemX Alliance, as well as collaboration and discussion with G. Pitner and colleagues at Taiwan Semiconductor Manufacturing Company (TSMC), M. Shulaker at MIT, and S. Mitra at Stanford. **Competing interests:** The authors declare no competing interests. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 15 June 2022; accepted 7 September 2022
10.1126/science.abp8278

REVIEW

Toward attoJoule switching energy in logic transistors

Suman Datta[1,2]*, Wriddhi Chakraborty[2], Marko Radosavljevic[3]

Advances in the theory of semiconductors in the 1930s in addition to the purification of germanium and silicon crystals in the 1940s enabled the point-contact junction transistor in 1947 and initiated the era of semiconductor electronics. Gordon Moore postulated 18 years later that the number of components in an integrated circuit would double every 1 to 2 years with associated reductions in cost per transistor. Transistor density doubling through scaling—the decrease of component sizes—with each new process node continues today, albeit at a slower pace compared with historical rates of scaling. Transistor scaling has resulted in exponential gain in performance and energy efficiency of integrated circuits, which transformed computing from mainframes to personal computers and from mobile computing to cloud computing. Innovations in new materials, transistor structures, and lithographic technologies will enable further scaling. Monolithic 3D integration, design technology co-optimization, alternative switching mechanisms, and cryogenic operation could enable further transistor scaling and improved energy efficiency in the foreseeable future.

In electronics, new inventions that reduce energy consumption and enhance integration capacity tend to become the dominant device platform. For example, the triode in a vacuum tube was used to amplify signals for almost half a century but engineers recognized drawbacks such as high power consumption resulting in high heat dissipation as well as overall lack of robustness. Energy consumption issues in telephone signal transmission drove researchers at Bell Laboratories to invent a solid-state semiconductor device as a more reliable and energy-efficient replacement for the vacuum tube amplifier. The bipolar point-contact was made from two isolated strips of gold foil on a plastic wedge, and these strips were pushed down to make contact with a slab of n-doped germanium. This was dubbed the transistor and was commercialized by their Western Electric subsidiary for radio transmission applications, although this configuration was eventually replaced by the easier-to-manufacture and more reliable silicon bipolar junction transistor.

For three decades, the silicon bipolar transistor would remain the device of choice in the design of both discrete and integrated circuits. The performance of the bipolar transistor improved with scaling but also led to increased self-heating and lower breakdown voltage. In the early 1980s, the increasing count of bipolar transistors on the integrated circuit reached an unacceptable level of power density. Managing power delivery and heat dissipation became a formidable challenge. For example, the iconic Cray-2 supercomputer of this era used dense packaging of multiple processors and fast bipolar transistor logic called emitter-coupled logic (*1*) to deliver two billion floating point operations per second (gigaFLOPs) of peak performance. The fast clock speed and dense packaging led to such high heat dissipation that the machine had to be continuously cooled through circulation of Fluorinert liquid (supplied by 3M) through the processor circuits. The visible heat exchange water tank that came with the liquid cooling system earned the Cray-2 the nickname "Bubbles."

During the late 1980s, there was a widespread design transition from bipolar to complementary metal oxide semiconductor (CMOS) transistor technology. CMOS, which had been in development since the 1960s and was favored by Japanese manufacturers, allowed for a much-needed reprieve from power challenges. Unlike bipolar transistors, complementary pull up and pull down transistor configuration of CMOS technology consumed negligible standby power while in a stable logical state, as it dissipated power only during state transition (*2*).

In its infancy CMOS was viewed by circuit designers as too slow compared with bipolar transistors for use in high-performance processors and was relegated to low-power and lightweight processing engines meant for data terminals, calculators, and avionics. Advances in high-resolution lithography techniques for transferring fine patterns to the integrated circuit and in ion implantation and dopant activation techniques for forming shallow source-drain junctions and tailoring channel doping profiles led to the design of highly scaled and fast-switching metal-oxide-silicon field-effect transistors (MOSFETs), which can have p- or n-type channels (PMOS and NMOS, respectively), CMOS used complementary and symmetrical pairs of these transistors. During the next four decades, transistor scaling would provide steady metronomic improvement for CMOS circuit performance. Gordon Moore's prediction that transistor count would double every 2 years (*3*) proved true (Fig. 1). Demonstrating clear advantages in energy efficiency and device density over bipolar transistors, MOSFET would become the primary choice for high-performance digital circuits for a broad range of technologies from general-purpose processors to domain-specific accelerators for cloud-enabled data centers and from desktop and mobile client computers to low-power embedded processors for wearable and internet-of-things applications.

The operating voltage of the MOSFETs decreased progressively as the lithographic dimensions shrank, leading to improved energy efficiency, faster speed, and lower cost per transistor. The lower energy consumption per switching event per transistor meant that the integrated circuit could accommodate more

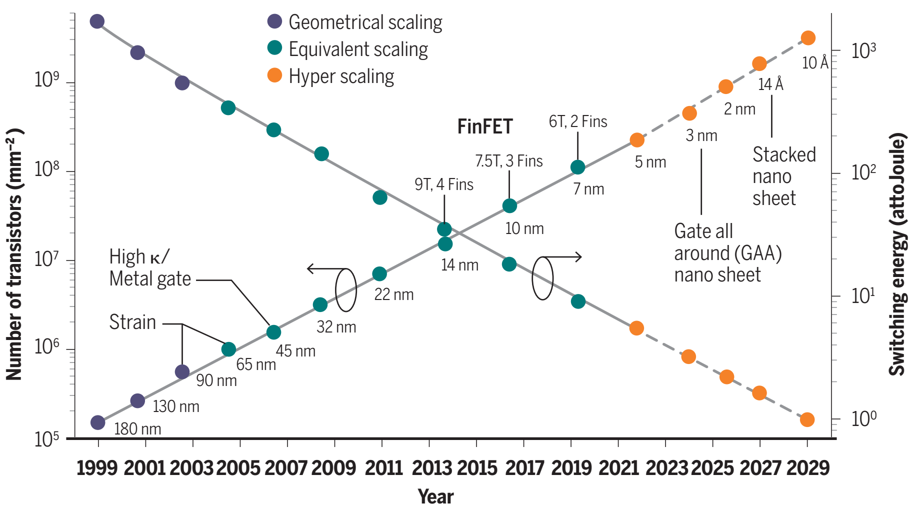


Fig. 1. Metronomic progress in CMOS transistor density and switching energy. Improvements in transistor density and switching energy are shown for geometrical, equivalent, and hyperscaling approaches. Examples for each approach are illustrated in Fig. 2.

[1]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA. [2]Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN, USA. [3]Components Research, Logic Technology Development, Intel Corporation, Hillsboro, OR, USA.
***Corresponding author. Email: sdatta68@gatech.edu**

logic gates that switched faster while maintaining the same operating power budget. The power dissipated is energy consumed per switching event divided by the temporal duration of the switching event. In theory, it is conceivable to keep the power consumption of the chip constant but in reality the power budget is often exceeded to meet the performance specification, especially for high-performance computing applications.

Complicating the matter further, modern processor chips use a heterogenous structure in which diverse functional blocks work with varying activity factors (activity factors refer to how often the transistors undergo state transitions), resulting in non-uniform dissipation of power on the chip. This spatial nonuniformity is accompanied by temporal nonuniformity as circuit blocks switch modes from idle to fully active at different points in time. Modern power-aware design techniques (e.g., power-down sleep modes, clock gating, dynamic voltage, and frequency scaling) exacerbate the spatiotemporal nonuniformity, creating "hot spots" on the chip. Failure to remove these hot spots in an expeditious manner not only affects chip performance but also causes errors in logic states, accelerates aging of the transistors, and in extreme cases results in premature failure of the integrated circuit.

Reflecting on the innovations that have enabled CMOS transistor scaling and elucidated ongoing efforts to improve energy efficiency, performance, and density of transistors in the future, this Review revisits past innovations and highlights future advances required for the transistor to continue its scaling trajectory and approach the milestone of attoJoule (aJ) energy consumption per switching event and density of over three billion transistors per square millimeter of silicon.

Geometric Scaling of Transistors following Dennard's Law

Robert Dennard and his colleagues at IBM T. J. Watson Research Center established the MOSFET scaling rules that result in simultaneous improvement in transistor density, switching speed, and power dissipation (*2*). Each new generation of CMOS process technology was expected to reduce the minimum transistor dimension from X in the current generation to 0.7X in the next, which then led to a 0.49X reduction in transistor area and thus a 2X increase in transistor density. The industry embraced Dennard's scaling and between 1974 (when the idea was introduced) and 2003 (when the 90 nm process, also called the 90 nm node, was introduced), the physical gate length of the transistor was successfully scaled from 1 μm to 35 nm. Additionally, the operating voltage was lowered from 4 V to 1.2 V, the transistor density increased from a few hundred to 6×10^5 transistors per square millimeter, and the switching energy per transistor fell below 1 picojoule (note that terms such as "90 nm process" refer to the design goals of the International Technology Roadmap for Semiconductors that assessed the research challenges for each size reduction). Operating voltage scaling was responsible for the reduction in the switching energy, whereas gate length scaling increased switching speed. During the era of geometric scaling, the processor clock frequency rose three orders of magnitude from 2 MHz in the case of the Intel 8008 processor (used to control traffic light signals) to 3 GHz in the case of the Intel Pentium D 64-bit dual-core processor (used to power client desktop computers).

It is implicit in Dennard's scaling law that the threshold voltage of the transistor, V_T, will scale proportionally with the operating voltage to ensure that there is sufficient voltage overdrive to provide high drive current. After three decades of transistor downsizing, V_T dropped so low that the subthreshold leakage current (the current that flows through the transistor even in its standby or off state) was now high enough to make standby power impose a constraint in addition to the dissipated dynamic power. A second implicit assumption in Dennard's scaling law is the continued scaling of the physical gate oxide thickness, which provides electrical insulation of the gate electrode from the current carrying transistor channel. After three decades of scaling, the thickness of the gate dielectric—which used a nitrogen-containing silicon dioxide—was reduced to 1.2 nm (4 monolayers thick) at the 90 nm node. With such a thin dielectric, the gate leakage current caused by direct quantum mechanical tunneling became a noticeable fraction of the standby power. These unacceptable levels of subthreshold and gate leakage current finally ended the era of geometric scaling and began that of equivalent scaling of transistors.

Equivalent Scaling of transistors

The era of equivalent scaling of transistors is defined by innovations in new materials and transistor structures in addition to dimensional scaling. In the absence of physical scaling of the SiO_2 gate oxide, transistor researchers pioneered three complementary approaches to improving transistor performance, energy efficiency, and scalability. Channel mobility was enhanced through the introduction of strain. The electrical gate oxide thickness was scaled by replacing SiO_2 with an alternative dielectric with high electrical permittivity κ. Planar, single-gate architectures were replaced with nonplanar, multigate structures to improve electrostatic integrity.

Strain engineering

Electron mobility in silicon increases with tensile strain as a result of splitting of the sixfold degenerate conduction band valleys, whereas the hole mobility increases with compressive strain as a result of the lifting of degeneracy between the light and heavy hole bands. A transistor structure with epitaxial source-drain regions can be designed to impart strain of both types to the channel (Fig. 2A). After the formation of gate stack and sidewall spacers as well as a silicon recess etch, selective heteroepitaxy can be performed to regrow a strained material in the source-drain regions adjacent to the channel. If the lattice spacing of this material is larger (or smaller) than silicon, uniaxial compressive (or tensile) strain is induced in the channel. The transistor designers at Intel engineered uniaxial compressive strain in the channel of PMOS transistors using selective silicon-germanium heteroepitaxy in the source-drain regions to boost hole mobility, and introduced uniaxial tensile strain in the channel of the NMOS transistors using a tensile silicon nitride capping layer to increase electron mobility (*4*). The widespread use of strained channel CMOS transistors in volume production at the 90 nm and 65 nm nodes heralded the era of equivalent scaling.

Gate dielectrics

Toward the end of the geometric scaling era, the gate dielectric SiO_2 layers were so thin that further decreases would effectively run out of atomic layers. There were two critical challenges to replacing the polycrystalline silicon (polySi) gate electrode and SiO_2 with polySi and higher permittivity dielectrics such as hafnium dioxide (HfO_2). First, HfO_2 and polySi were incompatible because charge sharing between the polySi and defect states within the HfO_2 created undesirable interface dipoles that led to unacceptably high V_T in the transistors. Second, polySi/HfO_2 transistors exhibited severely degraded channel mobility caused by scattering of the carriers in the channel by the soft optical (SO) phonons arising from the polarization of the HfO_2.

Metal gate electrodes proved effective in screening the high-κ SO phonons from coupling to the channel electrons and holes, whereas strain engineering in the channel provided further boosts in both electron and hole mobility (*5*, *6*). Transistor researchers successfully engineered the gate metal electrodes to exhibit the correct work functions (4.1 eV for NMOS and 5.1 eV for PMOS) using a replacement metal gate process flow. Also called the gate-last flow, the gate metal electrode and the high-κ dielectric are deposited after the high-temperature activation anneal associated with source-drain dopant activation. This sequence preserves the targeted threshold voltage, mobility, and reliability at scaled electrical oxide thicknesses. High-performance high-κ/metal gate CMOS with negligible gate leakage, targeted threshold voltages, low electrical oxide thickness, and

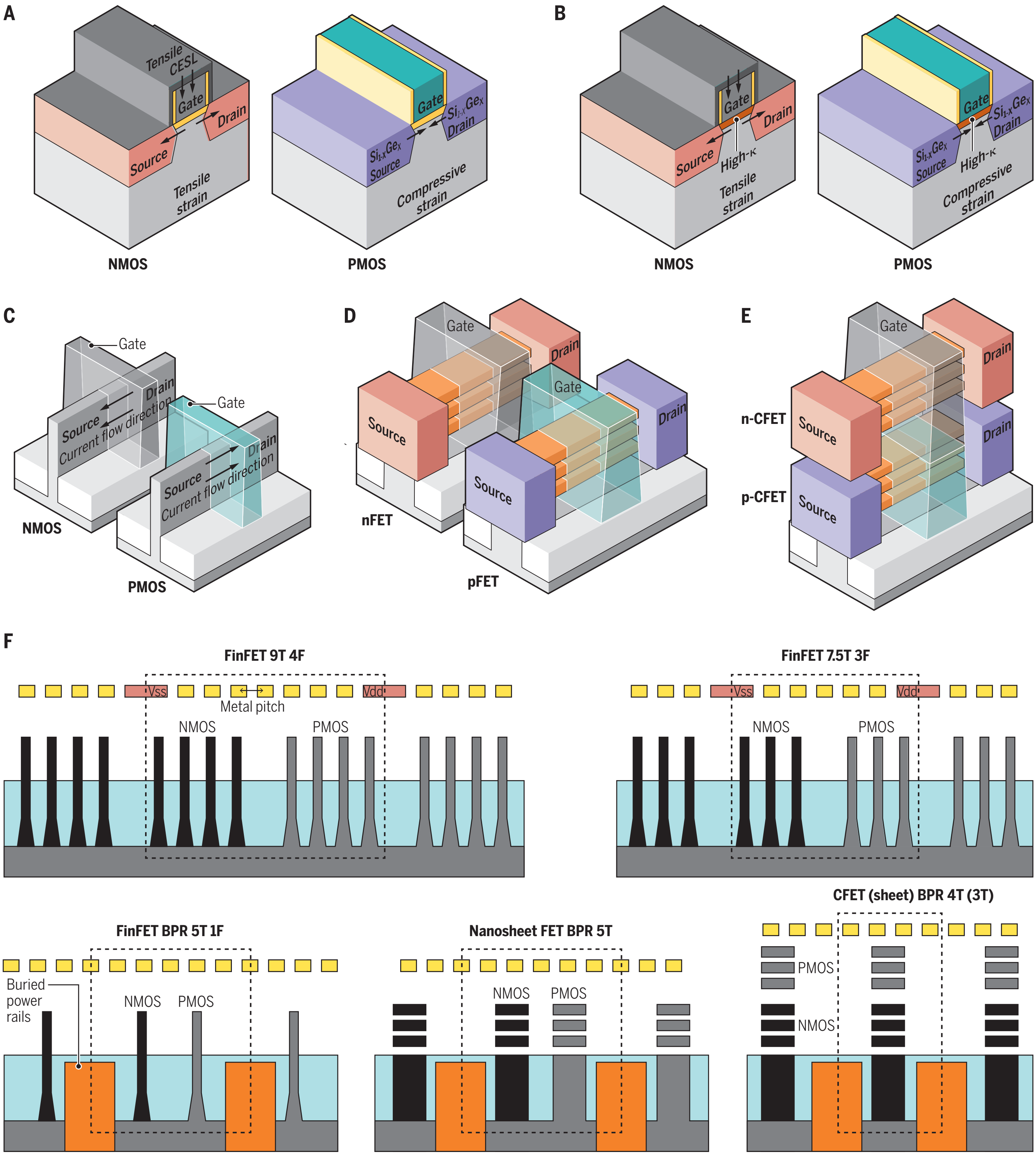


Fig. 2. Platforms enabling transistor improvements. Schematics show transistor architectures that have enabled geometrical and equivalent hyperscaling. (**A**) Strain engineering of silicon used vertical tensile strain to improve electron mobility in NMOS devices, whereas compressively strained selective silicon-germanium heteroepitaxy improved hole mobility in PMOS devices. (**B**) Metal gate electrodes and high κ gate dielectrics allowed scaling beyond the limits of SiO_2 as a gate dielectric. (**C**) The first nonplanar gate architecture, the FinFET, used vertical fins to create tri-gate structures with improved electrostatic confinement. (**D**) Electrostatic confinement will further improve with gate-all-around nanosheet transistors. (**E**) To further increase density, p-channel nanosheet transistors will be stacked on n-channel nanosheet transistors or vice versa. (**F**) DTCO provides further improvement in the power performance area for each node through codesign of the physical layout of the logic standard cell and specific technology choices. The design parameters include the overall transistor size as measured in track height T, fin height, spacing, and count (where F refers to the number of paired fins), the number of nanosheets stacked (which can be complementary in the CFET), and use of BPR.

channel mobilities for high-performance logic applications were developed with gate-last flow (Fig. 2B) (*7*). In the gate-first approach, developed by Sematech and the IBM-led Alliance, the high-k and metal gate stack goes through source-drain activation anneal and incorporates ultrathin capping layers—Al_2O_3 for the PMOS and LaO_x for the NMOS—to create dipoles that set the threshold voltage of the device (*8*). For low power or DRAM applications in which threshold voltage and electrical oxide thickness requirements are relaxed, gate-first became a viable and promising option for integrating a cost effective high-k and metal gate CMOS solution.

Nonplanar multigate architectures

High-κ dielectric and metal gate technology provided two generations of transistor scaling at the 45 and 32 nm nodes. Further scaling of the transistor gate length scaled with a single gate planar structure was no longer adequate to maintain electrostatic robustness and switch off the transistor effectively. The electrostatic integrity of the MOSFET is quantified by two key metrics: the subthreshold slope (SS) and the drain-induced barrier lowering (DIBL). Both SS and DIBL degraded as the channel length was further scaled beyond the 32 nm node. Researchers from the University of California at Berkeley and the Central Research Laboratory of Hitachi had previously proposed a nonplanar, self-aligned double-gate transistor structure called the FinFET (Fig. 2C). This design, which operated in the fully depleted mode and suppressed the short channel effect, provided a path for further scaling (*9*).

It would take more than a decade of effort for practitioners to combine the electrostatic benefit of the fully depleted FinFET transistor structure with high-κ gate dielectric, workfunction engineered metal gate electrodes and channel strain engineering in a triple-gate nonplanar geometry to demonstrate a high-performance transistor called the tri-gate transistor (*10*). The piezoresistive coefficients of silicon change their sign and magnitude as the function of the direction of the crystallographic planes (<100> versus <110>) for the top fin surface versus the fin sidewalls. The channel strain, high-k gate dielectric, low channel doping, gate electrode work functions, and selective epitaxial source-drain regions were co-optimized together with the fin dimensions (width and height) and the fin shape (trapezoidal versus rectangular) to demonstrate high-performance NMOS and PMOS tri-gate transistors.

The transition from single-gate planar transistors to triple-gate nonplanar transistors was a major milestone in the journey of transistor scaling, and FinFET was introduced by Intel for mainstream logic at the 22 nm node (*11*). FinFETs operated at lower supply voltages yet exhibited less leakage as a result of superior electrostatics compared with their planar counterpart, with a marked reduction in both dynamic and standby energy consumption. Because conduction occurs on all three surfaces of the fin, the device could drive more current through a given area of silicon than a planar transistor. The FinFET technology offered at the 5 nm node remains the workhorse of leading-edge logic technology (*12*).

Hyperscaling of transistors

As the contacted gate pitch (the smallest space between the gate electrodes of neighboring transistors) continues to scale and the transistor dimensions shrink, researchers are exploring structures beyond FinFETs to further improve electrostatics. They are also exploring the vertical direction (out of the silicon plane) for increasing the transistor density with gate-all-around (GAA) architectures where transistor channels, often called nanosheets (NSs) or nanoribbons, are completely surrounded by the gate dielectric and gate electrode (Fig. 3, D and E). This physical confinement provides even stronger electrostatic control of the charge carriers in the channel compared with the FinFET. This effect is illustrated in Fig. 3, A and B, where despite having the same critical channel parameters (fin width in case of FinFET and channel thickness in case of GAA), the NS structure provides better electrostatics and supports a shorter gate length at matched DIBL.

Design technology co-optimization (DTCO)—which is of paramount importance in the era of hyperscaling—optimizes together the standard logic cell and process technology features to improve performance, energy efficiency, density, and cost (Fig. 2F). Process technologists collaborate with standard cell designers from the very outset of each process node as they balance innovative design choices with process capability. In the early days of the technology transition between planar MOSFET to nonplanar FinFET, standard cells were designed with four fins per FinFET to deliver a drive strength equivalent to planar devices, resulting in standard cell tracks that are 9 tracks (9 T) high. In subsequent nodes, fin depopulation (three fins per FinFET) enabled technologists to reduce track height to 7.5 T and further maximize power, performance, and density gain. In the future, use of buried power rail (BPR) will lead to more aggressive track height reduction to 5 T.

In addition to scalability, NS transistors can be stacked several sheets high and provide the flexibility required in DTCO. In contrast to FinFETs where the width of the transistors comes in quanta of periphery of single fins (twice the fin height plus fin width), the width in NS transistors can be more readily tuned by lithography and optimized to provide higher

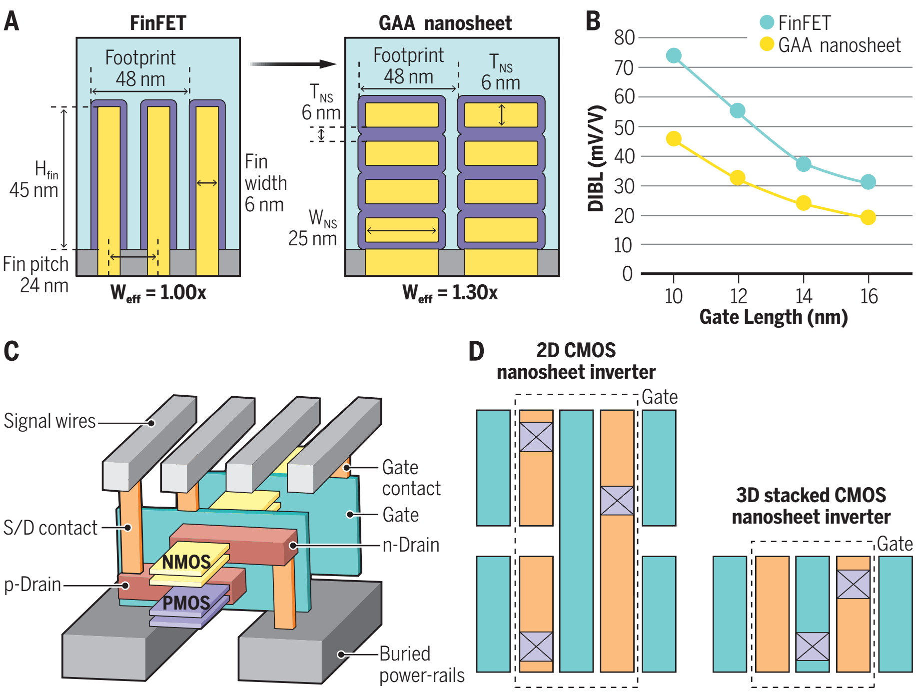


Fig. 3. Hyperscaling of transistors with stacked NSs. Compared with the FinFET design, a GAA transistor has (**A**) effectively higher effective gate width (W_{eff}) for the same geometric dimensions ("footprint") and (**B**) superior electrostatic gate control to enable further gate length scaling (*14*). Here, T_{sus} is nanosheet spacing, and T_{NS} and W_{ns} are the NS thickness and width, respectively. (**C**) Side-by-side stacked NS PMOS and NMOS transistors will be replaced by stacking NMOS on top of PMOS transistors [from (*37*) with permission]. (**D**) The additional density scaling resulting from vertical stacking is shown; V_{ss} and V_{cc} are source and drain supply voltage, respectively [from (*15*) with permission].

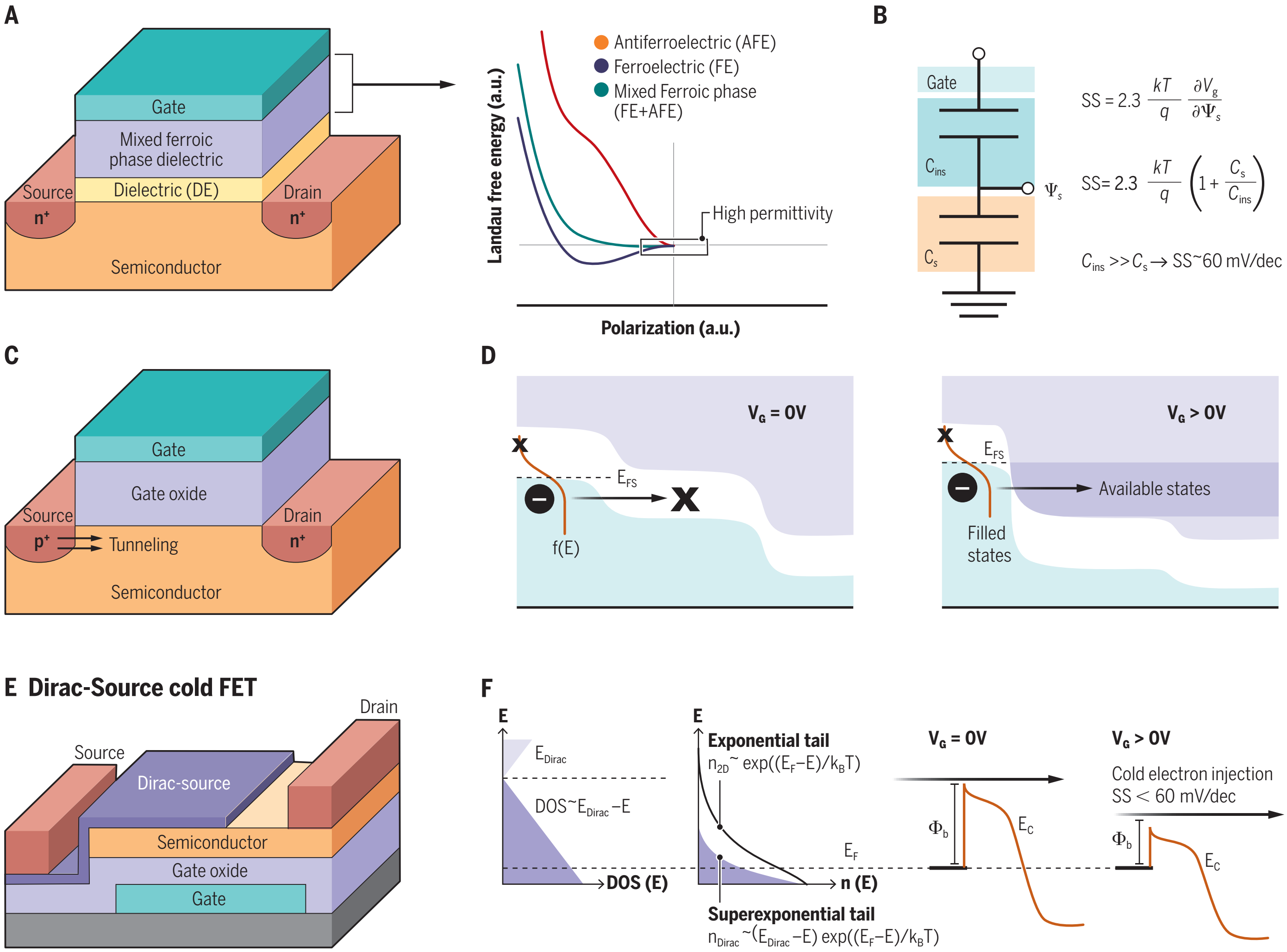


Fig. 4. Embodiments of unconventional transistors that can overcome the Boltzmann supply voltage limit of 500 mV. (**A** and **B**) Negative capacitance FETs use a mixed FE and AFE layer on top of the conventional dielectric to enhance the permittivity and maintain SS at ~ 60 mV per decade. (**C** and **D**) Tunnel FETs allow current to flow at positive gate voltages from immobile valence band states to mobile conduction band states [green region for (D)]. Electron tunneling occurs through states created in the bandgap when the bias voltages changes the band structure and narrows the gap between the valence and conduction bands. Here, $f(E)$ is Fermi function, the probability of an electron to occupy a state with energy E, as a function of the density of states (DOS), and E_{FS} is the energy of the highest filled state. (**E** and **F**) Dirac source cold FETs make use of semiconductors such as graphene that have a narrower super-exponential Boltzmann distribution of hot carriers that limits thermal injection. The number of 2D and Dirac carriers are n_{2D} and n_{Dirac}, respectively, E_F and E_C denote the Fermi energy and conduction energy band, respectively, and φ_B is the Schottky barrier.

overall current density per unit area (Fig. 3A). Transistors are designed with optimally designed widths to achieve a better balance of power, performance, and cost to meet circuit and product design specifications (*13*). All of these benefits are being harnessed now as major semiconductor manufacturers race to introduce GAA stacked NS transistors in the 3 nm nodes and beyond (*14*). Transistor density can be increased further with more advanced stacking techniques that place NMOS and PMOS NS transistors on top of each other rather than side-by-side (Fig. 3C). Nominally, by having two layers of transistors, density could be improved by as much as 50% (Fig. 3D) although more careful DTCO is needed that accounts for signal routing needs and leverages back-side power delivery options using BPR (*15*).

The concept of device stacking was proposed decades ago for co-integrating heterogeneous devices such as silicon (Si) and germanium (Ge) (*16*) or Si and compound semiconductors (such as GaAs) for additional functionality. However, for the purpose of density improvement the stacking must be done in an efficient manner. In the proposal shown in Fig. 3C, electrical coupling of stacked devices is facilitated by constructing interconnects on both the front and back side of the wafer (*17*). Stacking devices in this monolithic 3D fashion is an area of intense research worldwide. This architecture is referred to as complementary FET (or CFET) (*18*). In addition to being very promising, this density scaling approach does not depend on any specific transistor performance or scaling boosters, and as such system power-performance balance will come from design technology co-optimization effort.

Unconventional transistors operating at extremely low supply voltages

Density scaling can be enabled with GAA-stacked NS transistors along with design technology co-optimization approaches such as self-aligned source-drain contacts, contacts landing directly on active gates, buried supply rails, nanoscale through-silicon-vias, and direct backside source-drain contacts. However, the scaling of the supply voltage of operation of future transistors may slow down because of limits imposed

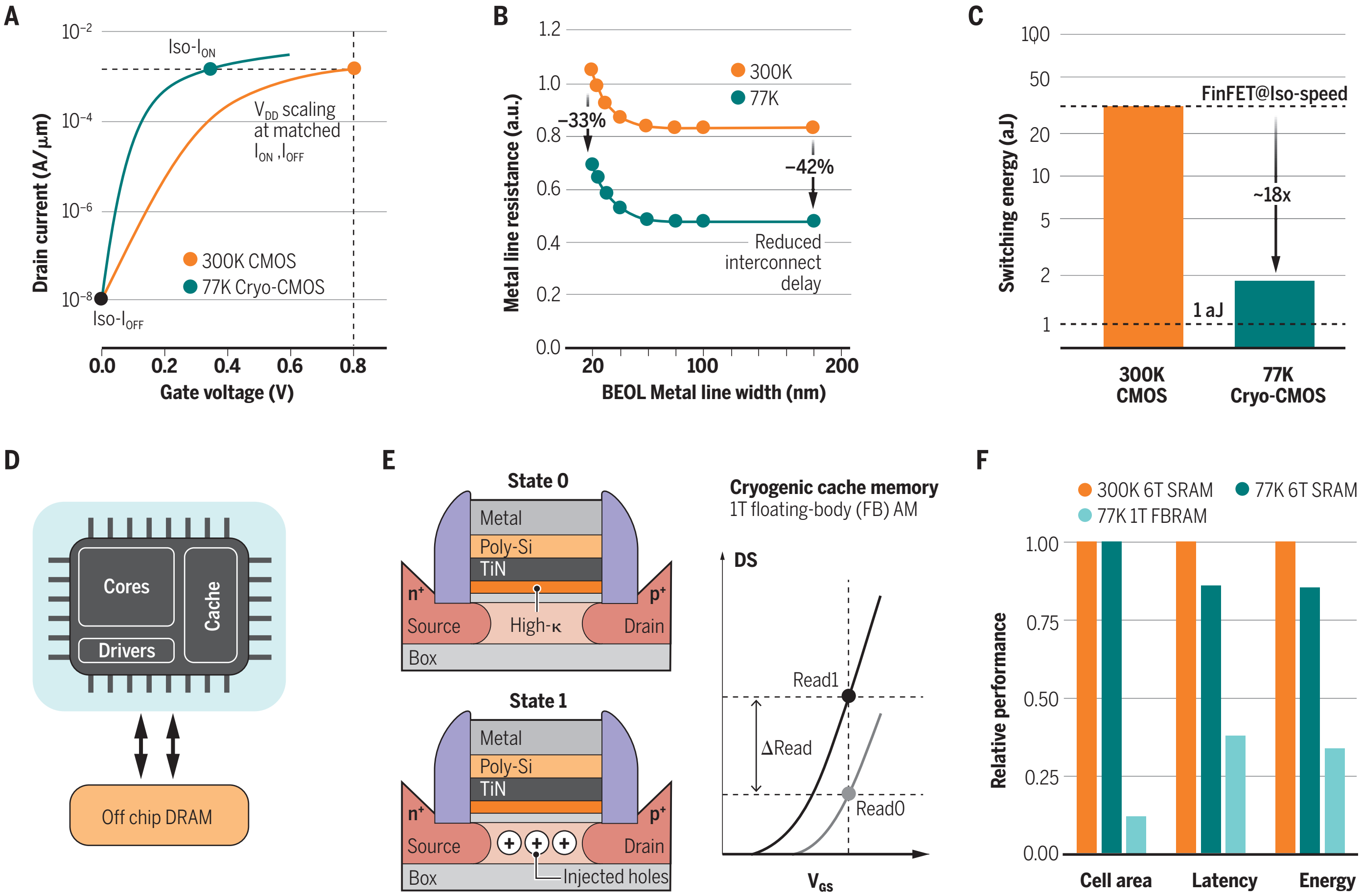


Fig. 5. Benefits of cryogenic operation. (**A**) Reduced back-end-of-line (BEOL) metal interconnect resistance compared with room temperature CMOS (*31*). (**B**) Cryogenic CMOS enables superior transistor performance attributed to key cryogenic boosters, such as higher channel mobility, steeper switching slope, and lower supply voltage operation. (**C**) Cryogenic CMOS shows potential for scaling of V_{DD} from 0.8 V at 300 K to 0.18 V at 77 K, hence approaching the aJ switching energy level. (**D** to **F**) Cryogenic floating-body DRAM provides low latency and high bandwidth access to high-capacity on-die memory to augment cryogenic SRAM, providing additional means to mitigate the memory bottleneck problem (*35*, *36*).

by the so-called "Boltzmann tyranny" as CMOS technology reaches its thermal and reliability limits. The MOSFET fundamentally operates as a barrier-controlled device where the gate electrode electrostatically modulates the potential barrier between the source and the channel to either impede or allow charge carrier injection into the channel.

However, the carriers follow Boltzmann statistics and there exists a high-energy tail in the distribution of carriers that enables injection of a small fraction of the charge over the barrier. Additionally, the fraction increases with temperature T and hence the MOSFET operates at the thermionic limit of k_BT/q (where k_B is Boltzmann's constant and q is the charge of an electron). The gate voltage must swing by at least 60 mV at room temperature to induce an order of magnitude change in source-to-drain current (the so-called SS voltage), and in practice imposes a lower limit on the minimum V_T of ~200 mV to keep the off-state leakage current at an acceptable level. The requirement of a minimum gate overdrive voltage of around 300 mV beyond the threshold voltage to deliver sufficient on-state current sets the minimum operating voltage of the MOSFET at 500 mV.

The inability to scale supply voltage below 500 mV will prevent CMOS technology from ever reaching aJ switching energies and set a practical limit of ~10 aJ at the proposed limits of hyperscaling. Doing otherwise will require revolutionary changes in the fundamental switching mechanisms of MOSFETs.

Negative capacitance

One such approach is negative capacitance FET (Fig. 4A), which boosts gate capacitance by stabilizing competing phases of ferroelectric (FE) and antiferroelectric (AFE) order in a HfO_2- ZrO_2 superlattice (*19*). Increasing gate capacitance allows the SS to remain at 60 mV per decade even at lower operating voltages. This stability results because SS scales as ($1 + C_S/C_{Ins}$), where C_S is the capacitance of the underlying semiconductor and C_{Ins} is the capacitance of the entire dielectric layer (Fig. 4B). However, with the addition of the mixed ferroic layer, C_{Ins} is now much larger than C_S and SS is now effectively constant. The conventional FE double-well energy landscape can be effectively flattened by the depolarization field arising from the AFE phase through electrostatic and elastic interactions. The energy landscape flattening has the effect of enhancing the permittivity of the overall stack and allows the use of a thinner electrical gate oxide of <6.5 Å. Resumption of gate dielectric thickness scaling in future generations of transistors will enhance control of the gate over the channel charge and allow a transistor operating voltage of <500 mV.

Tunneling

Tunnel FET (TFET) is another concept that replaces the conventional p-n junction formed by n-type sources and p-type channels in an NMOS transistor with a reverse biased interband tunnel junction using a p-type source and n-type channel (Fig. 4C). The switching slope in such a device is not limited to 60 mV per decade. Here, the thermal tail of the electrons in the

valence band of the p+ source is abruptly truncated by the energy gap of the source and can lead to steep switching characteristics as the TFET is turned on (Fig. 4D). A large indirect bandgap and relatively heavy carrier mass in silicon means that the drive current in silicon TFET is very low.

High-quality epitaxial growth of compound semiconductors and their incorporation into TFETs show promise from drive current perspective as a result of low bandgap and carrier mass. Staggered bandgap or even broken bandgap heterojunctions tend to have the highest TFET drive current (*20*, *21*). However, demonstration of TFETs with both high drive current and steep switching characteristics but operating voltages <500 mV remain elusive due to limitations of gate-dielectric quality for these newly introduced materials and difficulties in creating defect-free tunnel junctions (*22*).

Dirac source

The Dirac source FET is another candidate capable, in principle, of switching with subthermal slope and could potentially operate at low supply voltage (Fig. 4E). The charge carriers in the 2D gapless graphene source obey a linear energy dispersion relation mimicking massless relativistic Dirac particles. This property results in a narrower carrier density distribution around the Fermi level than that of the Boltzmann distribution of carriers in conventional MOSFETs (Fig. 4F). A Dirac source FET with a carbon nanotube channel was demonstrated with an average switching slope of 40 mV per decade over four decades of change in source-to-drain current at room temperature and a modest on-state current (*23*).

The intrinsic performance of future transistors benefits from new switching mechanisms discussed above. Still, source-drain (S/D) external resistance (R_{EXT}) may become a fundamental bottleneck for achieving maximized transistor performance. At the 5 nm node, contact resistance (R_{CON}) contributes to 40% of R_{EXT} with epitaxial S/D extension resistance (R_{EPI}) being the other limiting factor (*24*). This problem increases as the contact area continues to scale. Included among various approaches investigated for R_{CON} reduction is Schottky contact barrier modulation through insertion of an atomically thin interlayer (TiO_2) at the metal-Si interface (*25*), implementation of dual-layer metal stack (NiPt-Ti) (*26*), or even maximization of contact area through wrap around contact (*27*). By contrast, as dopant (P and B) concentration in Si S/D approaches the solid-solubility limit (~10^{20} atoms/cm^3), high activation of dopants through metastable Si:P alloying with 4 atomic % phosphorus concentration (*28*), solid phase epitaxially regrowth (SPER) of S/D through millisecond or nanosecond laser annealing (*29*) can lead to lower contact resistivity (ρ_c) as well as R_{EPI}.

Thermal management of transistors

Energy-efficient scaling of CMOS technology while still delivering on performance in the era of hyperscaling remains a challenge. New computational workloads such as training of large-scale deep neural networks involve a colossal amount of data movement. In such a situation, CMOS logic transistor performance needs to be balanced with high capacity, low latency (fast switching speeds), and high-bandwidth cache memory [near or on the central processing unit (CPU)] to avoid the so-called memory-wall bottleneck (*30*). Low-resistance metal line interconnects are also necessary to facilitate faster data movement between the logic and memory subsystems to maximize overall system performance (Fig. 5A) (*31*). Post-CMOS device technology with substantially improved energy-delay product (EDP) metrics remain to be developed (*32*).

Cryogenic CMOS—that is, CMOS operating at liquid nitrogen temperatures (77 K)—may provide a path toward scaling devices and reducing dynamic energy consumption to the aJ level. Cryogenic operation inherently enables steep SS CMOS logic transistors because SS scales linearly with temperature. Threshold voltage can now be further reduced by using gate electrode work-function engineering while maintaining iso-leakage as room temperature (300 K) CMOS (Fig. 5B). Improved carrier mobility and lower source/drain extrinsic resistance (*33*, *34*) at cryogenic temperatures further contributes to higher drive current and could enable aggressive supply voltage (V_{DD}) scaling.

> **"Energy-efficient scaling of CMOS technology while still delivering on performance in the era of hyperscaling remains a challenge. New computational workloads such as training of large-scale deep neural networks involve a colossal amount of data movement."**

Taking into account the energy cost of cooling to maintain the computing system at low temperatures, it is possible to demonstrate a net energy delay product (EDP) benefit. Low V_{DD} operation enables a pathway for equivalent oxide thickness scaling at cryogenic temperatures while maintaining iso-gate leakage, which opens up a broader design space for EDP optimization that takes into account device-to-device variation. A cryogenic-FinFET could potentially scale V_{DD} from 0.8 V at 300 K to 0.18 V at 77 K, which approaches the aJ switching energy level. The bulk resistivity of copper (Cu) interconnects also improves at low temperature (*31*) (Fig. 5C) and results in reduced interconnect delay and higher computing throughput.

Limited on-die cache memory capacity [a static random-access memory (SRAM) situated on the CPU itself] will still require a cryogenic processor core to access off-chip dynamic random-access memory (DRAM) for data-intensive workloads (Fig. 4D) and will limit performance and energy efficiency gains. Augmenting an on-die cryogenic SRAM cache with high-performance embedded memory will be critical to realize the full benefit of cryogenic CMOS technology. Cache memory is subcategorized into L1, L2, and L3 cache levels depending on its relative distance from the logic core. Because the L1 cache is placed near the processor and requires extremely low latency, cryogenic 6T-SRAM with 20% reduced latency and energy than 300 K SRAM (*35*, *36*) is a potential memory candidate.

Because the latency constraint can be somewhat relaxed for upper-level cache memory (L2 and L3), single transistor floating-body RAM (FBRAM) is being explored for cryogenic L2/L3 cache. Cryogenic one-transistor FBDRAM has been demonstrated with ~5-ns programming speed and 1.4 V programming voltage at 77 K (*36*). The floating-body effect in a single MOS transistor performs memory operations by injecting majority carriers into the "body" (the gate region) (Fig. 5E) to perform the write operation and are later expunged during an erase operation. The presence and absence of excess majority carriers in the floating body modulate the transistor source barrier and the drain current during the read operation (Fig. 5E). The low temperature operation enables higher transistor read current, suppresses Shockley-Read-Hall recombination rate and produces longer data retention time. The simpler 1T structure allows 8.33X higher capacity than 6T SRAM, along with 2.5X lower latency and 2.7X lower energy than 6T SRAM array (Fig. 5F) (*36*) and provides a superior core-to-cache balance, reduced power consumption, and faster computational throughput.

Outlook

During the next decade, CMOS transistor technology is poised to reach the unprecedented integration complexity of 2 billion transistors packed into one millimeter square of silicon with each transistor consuming only a few attoJoules of energy during switching operations. With an average projected switching time of less than a picosecond and an activity factor of a few percent, such an integrated circuit will consume only 20 watts of power. However, in order to achieve these goals, a multitude of innovations will be needed in materials, 3D transistor structures, interconnects, backside power delivery networks, and monolithic 3D integration

techniques, It will also require low latency access to on-die high-capacity memory beyond SRAM and fine-grained thermal management features.

REFERENCES AND NOTES

1. Fairchild Semiconductor, "CMOS, the Ideal Logic Family", Application Note 77 at the Wayback Machine, 1983.
2. R. H. Dennard *et al*., *IEEE J. Solid-State Circuits* **SC-9**, 256–268 (1974).
3. G. E. Moore, *Electronics* **38**, 114–117 (1965).
4. K. Mistry *et al*., *in Digest of Technical Papers at Symposium on VLSI Technology* (IEEE, 2004), pp. 50–51.
5. S. Datta *et al*., in *IEEE International Electron Device Meeting (IEDM) Technical Digest* (IEEE, 2003), pp. 653–656.
6. R. Chau *et al*., *IEEE Electron Device Lett.* **25**, 408–410 (2004).
7. K. Mistry *et al*., in *IEEE International Electron Devices Meeting (IEDM) Technical Digest* (IEEE, 2007), pp. 247–250.
8. K. Henson, in *IEEE International Electron Device Meeting (IEDM) Technical Digest* (IEEE, 2008), pp. 645–648.
9. D. Hisamoto *et al*., *IEEE Trans. Electron Dev.* **47**, 2320–2325 (2000).
10. J. Kavalieros *et al*., in *Digest of Technical Papers, IEEE Symposium on VLSI Technology* (IEEE, 2006), pp 62–63.
11. C. Auth *et al*., "A 22nm High Performance and Low-Power CMOS Technology Featuring Fully-Depleted Tri-Gate Transistors, Self-Aligned Contacts and High Density MIM Capacitors", Digest of Technical Papers, IEEE Symposium on VLSI Technology, 2012.
12. J. C. Liu *et al*., "A Reliability Enhanced 5nm CMOS Technology Featuring 5th Generation FinFET with Fully-Developed EUV and High Mobility Channel for Mobile SoC and High Performance Computing Application", IEEE International Electron Devices Meeting (IEDM) Technical Digest, 2020.
13. G. Bae *et al*., "3nm GAA Technology featuring Multi-Bridge-Channel FET for Low Power and High Performance Applications," 2018 IEEE International Electron Devices Meeting (IEDM), 2018, pp. 28.7.1-28.7.4, doi: 10.1109/IEDM.2018.8614629.
14. J. Zhang *et al*., "Full Bottom Dielectric Isolation to Enable Stacked Nanosheet Transistor for Low Power and High Performance Applications," 2019 IEEE International Electron Devices Meeting (IEDM), 2019, pp. 11.6.1-11.6.4, doi: 10.1109/IEDM19573.2019.8993490.
15. C.-Y. Huang *et al*., in *IEEE International Electron Devices Meeting (IEDM) Technical Digest* (IEEE, 2020), pp. 20.6.1–20.6.4.
16. S. Subramanian *et al*., in *Digest of Technical Papers, IEEE Symposium on VLSI Technology* (IEEE, 2020), pp. 1–2.
17. W. Rachmady *et al*., in *IEEE International Electron Devices Meeting (IEDM) Technical Digest*, (IEEE, 2019), pp. 29.7.1–29.7.4.
18. J. Ryckaert *et al*., in *2018 IEEE Symposium on VLSI Technology* (IEEE, 2018), pp. 141–142.
19. S. Cheema *et al*., *Nature* **604**, 65–71 (2022).
20. D. K. Mohata *et al*., in *IEEE Electron Devices Meeting (IEDM) Technical Digest* (IEEE, 2011), pp. 33.5.1–33.5.4.
21. G. Dewey *et al*., in *IEEE Electron Devices Meeting (IEDM) Technical Digest* (IEEE, 2011), pp. 33.6.1–33.6.4.
22. Y. Qiu, R. Wang, Q. Huang, R. Huang, *IEEE Trans. Electron Dev.* **61**, 1284–1291 (2014).
23. C. Qiu *et al*., *Science* **361**, 387–392 (2018).
24. N. Breil, in *2021 IEEE International Interconnect Technology Conference (IITC)* (IEEE, 2021), pp. 1–3.
25. A. Agarwal *et al*., *Appl. Phys. Lett.* **104**, 112101 (2014).
26. P. Adusumilli *et al*., in *2016 IEEE Symposium on VLSI Technology*, (IEEE, 2016), pp. 1–2
27. S.-A. Chew *et al*., in *2017 IEEE International Interconnect Technology Conference (IITC),* (IEEE, 2017), pp. 1–3.
28. C.-N. Ni *et al*., in *2016 International Symposium on VLSI Technology, Systems and Application (VLSI-TSA)*, (IEEE, 2016), pp. 1–2.
29. Y. R. Yang *et al*., in 2016 *EEE Symposium on VLSI Technology*, (IEEE, 2016), pp. 1–2.
30. H. L. Chiang *et al*., in *2021 Symposium on VLSI Technology*, (IEEE, 2021), pp. 1–2.
31. H. L. Chiang *et al*., in *2020 IEEE Symposium on VLSI Technology*, (IEEE, 2020), pp. 1–2.
32. C. Pan, A. Naeemi, *IEEE J. Explor. Solid-State Comput. Devices Circuits* **3**, 101–110 (2017).
33. W. Chakraborty *et al*., in *Digest of Technical Papers, IEEE Symposium on VLSI Technology*, (IEEE, 2021), pp. 1–2.
34. W. Chakraborty *et al*., in *IEEE International Electron Devices Meeting (IEDM) Technical Digest*, (IEEE, 2019), pp. 39.4.1–39.4.4.
35. W. Chakraborty *et al*., in *IEEE International Electron Devices Meeting (IEDM) Technical Digest*, (IEEE, 2021), pp. 40.1.1–40.1.4.
36. W. Chakraborty *et al*., in *Digest of Technical Papers, IEEE Symposium on VLSI Technology and Circuits*, (IEEE, 2022), pp. 302–303.
37. L. Liebmann *et al*., *in 2021 IEEE International Electron Devices Meeting* (IEDM), (IEEE, 2021), pp. 3.1.1–3.1.4.

ACKNOWLEDGMENTS

Funding: W.C. and S.D. would like to acknowledge the support by the Defense Advanced Research Projects Agency (DARPA) Low Temperature Logic Technology (LTLT) Program. **Competing interests:** The authors declare no competing interests. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

Submitted 6 September 2022; accepted 12 October 2022
10.1126/science.ade7656

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**



The addition of metal nanoclusters enables 3D printing of intricate and durable polymer lattices.

PHOTO: LI *ET AL.*

3D PRINTING

Metal nanoclusters improve printing

To enhance the properties of two-photon three-dimensional (3D)–printed polymeric structures, photoinitiators are added to increase the formation of radicals and to improve the polymerization process. However, each polymer may require a specific initiator, and the organic molecules typically used do not enhance the properties of the finished product. Li *et al*. report an improvement in two-photon 3D printing in which atomic metal clusters are used as two-photon absorbers to enhance the mechanical properties of the printed structures. The authors reveal the formation of a complex architecture with tunable and hierarchical porosity showing high mechanical strength and stability. —MSL

Science, abo6997, this issue p. 768

SOLAR CELLS

Staying distributed with selenophene

In perovskite solar cells, mixtures of cations and anions can be used to tune band gaps and increase stability, but these mixtures are prone to unwanted segregation that degrades performance. Bai *et al*. found that segregation can be slowed by creating an initially homogeneous distribution of cesium and formamidinium cations in the colloidal film precursors. The additive selenophene maintained homogeneity during film processing and device operation for both single- and mixed-halide perovskites. Iodide-based solar cells could retain 90% of their efficiency under illumination after 3000 hours at 45°C. —PDS

Science, abn3148, this issue p. 747

PLANT SCIENCE

Lateral root development

Plant roots are most effective in pockets of soil containing the water that the plant needs. Water is not uniformly distributed throughout the soil, and the xerobranching response, in which lateral root formation is suppressed, ensures that roots are not uniformly distributed either. Mehra *et al*. show what happens as plant roots grow into and through dry pockets in soil. In moist conditions, water flows from the root surface into the phloem, but in a dry spot, water flows instead from the phloem out into the root tissues. This reverse flow carries along the phloem-derived hormone abscisic acid, which closes off intercellular pores, blocking the ability of the hormone auxin to initiate lateral root development. —PJH

Science, add3771, this issue p. 762

CHEMICAL PHYSICS

Spectroscopy for polaritonic chemistry

There is currently considerable interest in understanding the effect of vibrational strong coupling (VSC), a promising route to manipulating chemical dynamics in condensed phases. Previous studies of VSC-modified chemistry used traditional chemical kinetics tools that cannot properly address ultrafast vibrational processes in their natural time scales. Using ultrafast two-dimensional infrared spectroscopy, Chen *et al*. showed that polaritons (delocalized superpositions of vibrations and electromagnetic cavity modes) can switch the rates of two vibrational energy pathways of a metal carbonyl compound under VSC, making intramolecular vibrational redistribution more favorable

over pseudorotation (the opposite is true outside of a cavity) (see the Perspective by Chuntonov). The authors also clarified the role of dark modes in VSC, a longtime question that has been heavily debated but had lacked direct experimental evidence. —YS

Science, add0276, this issue p. 790; see also ade9815, p. 712

QUANTUM PHYSICS

Tough edges

The dynamics of quantum many-body systems can be profoundly affected by their interaction with the environment. This includes systems that have topological protection from certain kinds of perturbations due to symmetry. Mi *et al*. studied the interplay between symmetry and noise using a chain of 47 superconducting qubits. They implemented a periodically driven transverse Ising spin model, and found that the system's edge modes were surprisingly resilient to some types of symmetry-breaking noise. —JS

Science, abq5769, this issue p. 785

NEURODEGENERATION

Protecting the brain from prions

Various neurodegenerative diseases are caused by prions or prion-like misfolded proteins that aggregate and propagate through the brain. Because loss of acetylcholine signaling is associated with cognitive deficits in patients, Dwomoh *et al*. investigated the effect of enhancing acetylcholine signaling in a mouse model of prion disease. Systemic administration of a positive allosteric modulator targeting the M_1 acetylcholine receptor reduced the levels of prion-induced molecular markers in the hippocampus, restored various cognitive functions, and slowed disease progression in the mice. —LKF

Sci. Signal. **15**, eabm3720 (2022).

CORONAVIRUS

Casting a wide NET on long COVID

A substantial number of individuals who recover from COVID-19 still present with long-term sequelae. George *et al.* followed individuals after recovery from severe COVID-19 to identify features that distinguished those who had evidence of long-term pulmonary sequelae from those who made a full recovery. They found that a neutrophil-associated inflammatory phenotype was apparent in those who had persistent pulmonary symptoms, and evidence of neutrophil extracellular traps, or NETs, was found in the blood of these individuals. These data highlight a potential role for neutrophils in pulmonary long COVID. —CSM

Sci. Transl. Med. **14**, eabo5795 (2022).

HUMAN GENOMICS

Nonrandom selection

Many studies have examined correlations between complex traits, assuming that correlations implied a genetic connection even when there was no clear biological overlap. It had been proposed that overlapping genes with pleiotropic effects contribute to multiple different psychiatric disorders or even across disease categories such as psychiatric and metabolic conditions. By combining analysis of phenotype data from two large, population-based cohorts with in silico simulations, Border *et al*. demonstrated that many correlations between human traits can be explained instead by cross-trait assortative mating, which is an individual's tendency to choose a mate with specific characteristics that have no genetic relationship (see the Perspective by Grotzinger and Keller). —YN

Science, abo2059, this issue p. 754; see also ade8002, p. 709

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**



HOST AND PATHOGEN

Ex vivo model ticks all the right boxes

Lyme disease, which is caused by *Borrelia burgdorferi*, is transmitted by ticks of the genus *Ixodes*. Although tick saliva is generally known to be immunosuppressive, the effects of tick feeding and the accompanying tissue damage on local and systemic immunity are not well understood. Strobl *et al.* recruited individuals with recent tick bite history and took skin biopsies from both the site of the bite and a control area. Neutrophils, B cells, and T cells (particularly tissue-resident memory T cells) were increased at the site of the bite, whereas other immune cell populations such as Langerhans cells were decreased. Cytokine production by T cells was impaired, and innate lymphoid cells at bite sites were also decreased. The immunomodulatory effects could be replicated by injecting tick saliva, along with *Borrelia* spirochetes, into human skin explants, providing a potential model to study human-tick interactions ex vivo. —STS

J. Clin. Invest. 10.1172/JCI161188 (2022).

VIROLOGY

Some viruses mix it up

Respiratory viruses such as influenza virus (IAV) and respiratory syncytial virus (RSV) are common among children and may cause coinfections, sometimes with exacerbated symptoms. Such infections are likely to have different dynamics from single infections. Haney *et al.* examined IAV and RSV coinfections in human lung cells in vitro and observed hybrid virus particles with surface glycoproteins and ribonucleoproteins from both viruses. Although RSV tends to be at a disadvantage in

PHOTO: FRANS LEMMENS/ALAMY STOCK PHOTO

coinfections with IAV, the hybrids could apparently evade anti-influenza antibodies by using the RSV fusion glycoprotein to get into cells from which IAV receptors had been removed. Proteins from both viruses were found to colocalize on the apical side of bronchial epithelial cells. The authors did not test whether these hybrid viruses are transmissible between animals in vivo. —CA

Nat. Microbiol. **7**, 1879 (2022).



BEHAVIOR
Too hot to shoal

Climate change is altering a wide array of environmental conditions, and we know little about how these changes may affect various traits such as social behavior. In fishes, research has shown that sociability can be lower in warmer temperatures, a possible result of selection for reduced disease transmission or resource competition. To test for a direct influence of warmer waters, Pilakouta *et al.* took advantage of a unique natural system in which adjacent and overlapping populations of three-spine stickleback live in either cold pools or those warmed by geothermal outflow. They found that the fish from isolated cold pools were more sociable than those from warm pools, but there was no evidence of plastic responses to water temperature in adult fish. Common garden rearing experiments, however, revealed that sociability was heritable and that the fish reared in warmer waters were less sociable, suggesting that temperature changes could alter fundamental behaviors. —SNV

Glob. Chang. Biol. 10.1111/gcb.16451 (2022).

Three-spined sticklebacks tend to form shoals in cool water, but not in warm conditions.

METALLOPROTEINS
A design that bucks the trend

Although metal sites in proteins can be somewhat selective, they often follow a predictable trend in which copper ions are the strongest binders. Choi *et al.* redesigned an artificial metal-binding protein such that two molecules were linked by a disulfide bond but were still slightly flexible and able to adopt a variety of conformations. This construct bound zinc, nickel, and copper ions and displayed selectivity for zinc in competition experiments, possibly enabled by binding three ions instead of one. Structural and biophysical analysis of the binding mode reveals how tetrahedral coordination geometry guides metal selectivity in this system. —MAF

J. Am. Chem. Soc. **144**, 18090 (2022).

ORGANIC CHEMISTRY
Trapping radicals

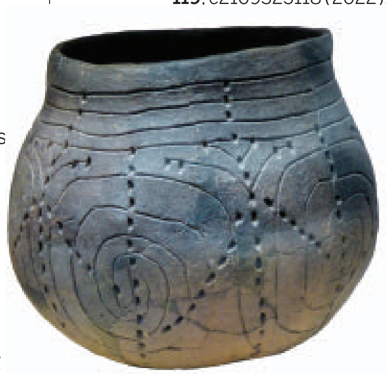
Many chemical reactions proceed through radicals, highly reactive intermediates bearing an unpaired electron. Because they are often short-lived, radicals can be hard to detect, but verifying their presence is important to optimize conditions and, in some cases, to avoid hazardous unanticipated chain reactions. Williams *et al.* report a trapping reagent that reacts efficiently with a wide variety of radicals by the displacement of a nitroxide leaving group on a carbon adjacent to an olefin. Its reactivity toward non-carbon-centered radicals is a particular improvement over prior probe compounds. —JSY

J. Am. Chem. Soc. **144**, 15969 (2022).

ANTHROPOLOGY
Updating carbon dating

Lactose tolerance is a classic example of selection in humans that substantially expanded available food sources with the advent of agriculture. However, timing the rise of dairy consumption is difficult from archaeological sources alone. Casanova *et al.* used a new method of carbon dating directly on the dairy fat residu found in pottery. They examined pottery vessels from across central Europe and found locations where there was no evidence of dairy consumption and others with low levels. Their dating estimates are earlier than those of previous studies and indicate that dairy consumption was adopted more widely and less gradually than was previously believed. —CNS

Proc. Natl. Acad. Sci. U.S.A. **119**, e2109325118 (2022).



An example of a Linearbandkeramik vessel, in which traces of dairy fat enabled carbon dating of Early Neolithic farming activities in Europe.

HISTORY OF SCIENCE
A fragment of Hipparchus' catalog

The second-century BCE Hellenistic astronomer Hipparchus is known to have produced a star catalog, but no copy of it has survived. Gysembergh *et al.* have reexamined previously recorded multispectral imaging of a palimpsest, a parchment that was incompletely erased before being reused for a different text. They identified about 100 words of faint Greek writing describing the constellation Corona Borealis and four of its brightest stars. The coordinate system and positions identify it as a copied part of Hipparchus' catalog. Although brief, the fragment is sufficient to demonstrate that Hipparchus' catalog was more precise than the one produced by Claudius Ptolemy three centuries later. —KTS

J. Hist. Astron. 10.1177/00218286221128289 (2022).

PHOTO: WOLFGANG SAUBER/ CC BY SA 3.0

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

CORONAVIRUS

Measuring menstruation after vaccines

Menstrual changes, particularly longer cycles and heavier bleeding, have been reported in association with various vaccines, including those for COVID-19. These observations have led to widespread misinformation that resulted in COVID-19 vaccine hesitancy among young women. Various studies have since found that although menstruation can be affected by COVID-19 vaccination, these changes are transient and resolve within a few months. In a Perspective, Male discusses studies looking at how COVID-19 vaccination and infection affect menstruation and the possible underlying mechanisms. These effects are not routinely monitored in vaccine trials, and this is a missed opportunity that could reveal feedback between the immune system and the female reproductive system. —GKA

Science, ade1051, this issue p. 704

GEOLOGY

Defining the Anthropocene epoch

A new geological epoch is upon us. Moving beyond the Holocene, we are now in the Anthropocene, which is proposed to have begun around the mid-20th century. The definition of the Anthropocene involves the identification of specific markers in sediment layers that capture the influence of increased human population on the planet. In a Perspective, Waters and Turner discuss 12 proposed sites that could be used as the reference for defining the beginning of the Anthropocene. Cores from these sites contain various markers that allow precise dating to fix the onset of the Anthropocene. These candidate sites will be voted on in late 2022. If a site is eventually selected, it will be used to help focus studies on how human activity has affected the planet. —GKA

Science, ade2310, this issue p. 706

DEVELOPMENTAL BIOLOGY

Rethinking Hippo signaling

Mutations that cause dysregulation of the Hippo signaling pathway are known to cause excessive growth of organs, which has led many researchers to think of this pathway as a master regulator of organ growth. Studying fruit fly eye discs and mouse livers, Kowalczyk *et al*. found instead that Hippo signaling does not instruct normal growth. The Hippo-induced overgrowth phenotypes appear to be caused by Hippo signaling activating abnormal genetic programs. These findings challenge a long-standing idea about the role of Hippo signaling in organ growth and suggest the need to re-evaluate our understanding of its function in other contexts, such as in cancer and regeneration. —SMH and BAP

Science, abg3679, this issue p. 744

HUMAN FERTILITY

Organizing for meiotic success

Proper organization of the microtubules is vital for ensuring that daughter cells end up with the appropriate complement of chromosomes in both mitosis and meiosis. In somatic cells undergoing mitosis, this task is performed by centrosomes. By contrast, meiosis does not involve centrosomes in many animal species, but the specific methods of organizing the microtubules differ between animals. In particular, the mechanism of spindle organization in human oocytes has not been previously understood. Wu *et al*. detected a protein structure that they named the human oocyte microtubule organizing center and identified several of its constituent proteins. They then showed that mutations in one of these proteins are responsible for the clinical infertility associated with oocyte maturation arrest in human patients. —YN

Science, abq7361, this issue p. 745

CANCER

FMRP and tumor immunity

Many tumors have developed mechanisms rendering them resistant to attack and destruction by the immune system. Zeng *et al*. report that fragile X mental retardation protein (FMRP) is highly expressed in human cancers, and they propose that it is involved in antitumor immunity. FMRP is best known as an RNA-binding protein that regulates the stability and translation of neuronal RNAs. By genetically inactivating the FMRP gene in mouse cancer cells, the researchers found that FMRP-deficient tumors had reduced growth and were more susceptible to attack by T lymphocytes. Tumor cells lacking FMRP showed remodeling of the tumor microenvironment, macrophage polarization, and up-regulation of the chemokines involved in effector $CD8^+$ T cell recruitment. —PNK

Science, abl7207, this issue p. 746

ORGANIC CHEMISTRY

Activating pyridine's 3-position

Numerous pharmaceutical compounds contain aromatic heterocycle components such as pyridine. Chemists therefore prize reactions that selectively append substituents to particular sites on these rings. Boyle *et al*. report that upon opening pyridine rings to form linear imines, the distinct characteristics of these intermediates strongly favored halogenation at the carbon in the otherwise unreactive 3-position. Recyclization then yielded a selectively halogenated ring. A complementary approach by Cao *et al*. relies on a reversible reaction of the pyridines with an alkyne and ester, disrupting the electronic structure to activate the 3-position for halogenation as well as trifluoromethylation (see the Perspective by Joo). —JSY

Science, add8980 and ade6029, this issue, pp. 773 and 779; see also ade5501, p. 710

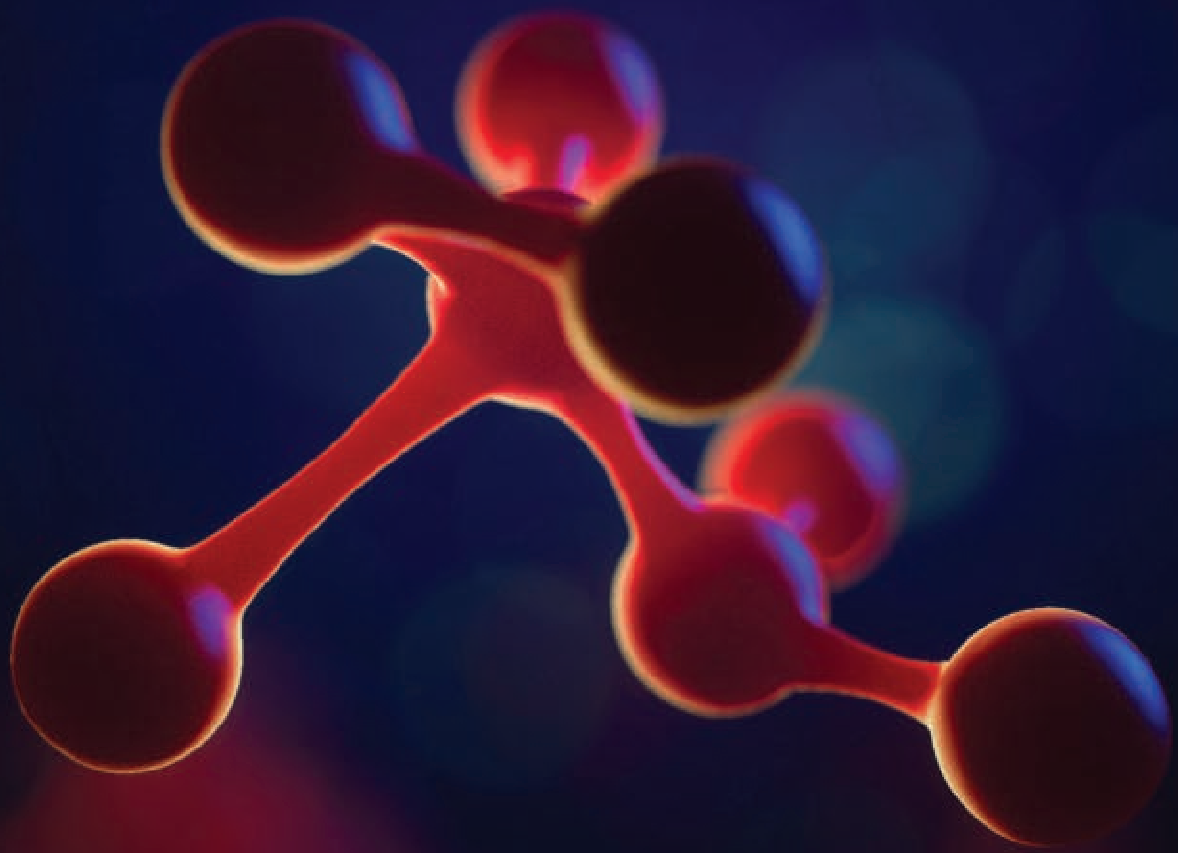
SKIN INFLAMMATION

Sustenance for skin-resident T_H17 cells

The T helper 17 (T_H17) cell subset provides host protection from extracellular pathogens but also contributes to autoimmune diseases. Interleukin-23 (IL-23) promotes cutaneous T_H17 responses, but the source of the IL-23 that supports skin-resident memory T_H17 cells has been unclear. Whitley *et al.* characterized the effects of blocking IL-23–mediated signaling in mice on T_H17 function in defending against skin infection by *Candida albicans* and in promoting inflammatory dermatitis. Dermal myeloid cells were identified as the critical source of IL-23 needed to sustain skin-resident memory T_H17 cells. These findings provide a deeper understanding of how blocking antibodies targeting either IL-23 or its receptor can hold aberrant cutaneous T_H17 responses in check, providing a durable therapeutic benefit to patients with T_H17-mediated inflammatory skin diseases such as psoriasis. —IRW

Sci. Immunol. **7**, eabq3254 (2022).

Science
JOURNALS AAAS



Publish your research in the *Science* family of journals

The *Science* family of journals (*Science, Science Advances, Science Immunology, Science Robotics, Science Signaling*, and *Science Translational Medicine)* are among the most highly-regarded journals in the world for quality and selectivity. Our peer-reviewed journals are committed to publishing cutting-edge research, incisive scientific commentary, and insights on what's important to the scientific world at the highest standards.

Submit your research today!

Learn more at **Science.org/journals**

RESEARCH ARTICLE SUMMARY

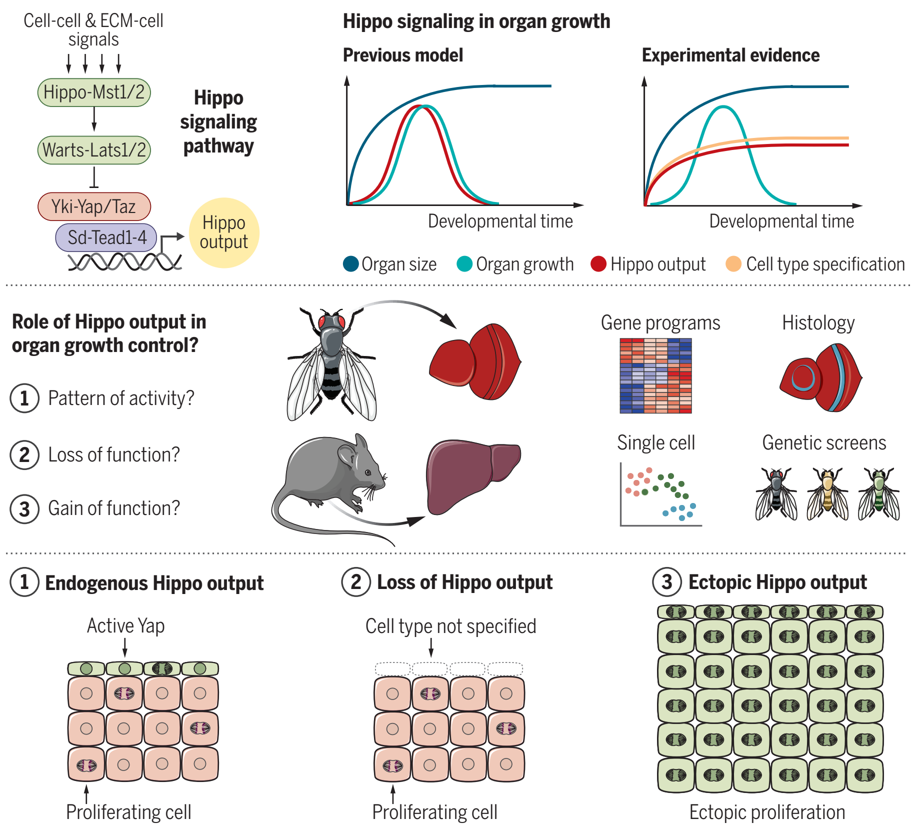
DEVELOPMENTAL BIOLOGY

Hippo signaling instructs ectopic but not normal organ growth

W. Kowalczyk†, L. Romanelli†, M. Atkins, H. Hillen, C. Bravo González-Blas, J. Jacobs, J. Xie, S. Soheily, E. Verboven, I. M. Moya, S. Verhulst, M. de Waegeneer, L. Sansores-Garcia, L. van Huffel, R. L. Johnson, L. A. van Grunsven, S. Aerts, G. Halder*

INTRODUCTION: Accurate growth control is fundamental to ensure proper organ size and animal health. Many signaling pathways can regulate cell proliferation and apoptosis, which in turn determine the size of an organ. However, mechanisms that can sense when a developing organ has reached its proper size and instruct the cessation of further growth remain elusive. Such mechanisms may also orchestrate regeneration and induce tumorigenesis when defective. In this context, the Hippo signaling pathway has attracted much attention, because experimental hyperactivation of its downstream effectors, the transcriptional coactivators Yap/Taz/Yki, can drive cell proliferation and cause substantial organ overgrowth and tumor development.

RATIONALE: The Hippo pathway integrates diverse biochemical and mechanical cell-cell and cell–extracellular matrix signals to regulate gene expression through Yap/Taz/Yki. In a current model of growth control, the activity of Yap/Taz/Yki directs the right amount of cell proliferation during development such that they are active when organs grow and inactivated when they stop growing. We reasoned that a signaling pathway functioning as a central regulator of organ growth should fulfil three requirements: (i) normal growth does not start or terminate properly without its input, (ii) its activity correlates with the spatial and temporal pattern of cell proliferation, and (iii) when hyperactivated, it produces larger but otherwise normal organs. We thus investigated whether the Hippo pathway fits these criteria for an organ growth control mechanism.



Functional analysis of the Hippo pathway in organ growth control. The regulation of the Hippo signaling pathway has been thought to determine the amount of organ growth during development. However, analysis of the endogenous activity and effects of genetic manipulations of the pathway in the developing mouse liver and *Drosophila* eye show that Hippo signaling does not instruct normal organ size. In the Hippo pathway schematic (top left), *Drosophila* proteins are shown on the left and mouse proteins on the right, separated by a hyphen.

RESULTS: To address these hypotheses, we studied two commonly used models for organ growth control, the *Drosophila* eye and the mouse liver, in which we examined the effects of loss of Hippo signaling output on growth, its pattern of activity during normal development, and the genetic program induced by the experimental hyperactivation of Yap/Taz/Yki.

Our data show that Hippo pathway output is not required for the developmental proliferation of liver precursor and *Drosophila* retinal cells. Furthermore, the transcriptional activity of Yap/Taz/Yki did not correlate with cell proliferation during development, and their experimental hyperactivation in hepatocytes and *Drosophila* retinal cells, which caused overgrowth, did not reactivate progenitor programs. Rather, Yap/Taz/Yki were active and required in specific cell types, namely *Drosophila* squamous peripodial epithelial cells and mouse cholangiocytes, and their hyperactivation induced aberrant gene expression profiles that partly resembled the normal programs of these cells. Finally, a functional screen in *Drosophila* identified several Hippo pathway target genes that were required for ectopic overgrowth but not for normal growth.

CONCLUSION: Our data show that Hippo signaling does not instruct normal organ growth, and that the classic Hippo mutant overgrowth phenotypes represent Yap/Taz/Yki gain-of-function situations that involve ectopic expression of genes that are normally expressed in Hippo-dependent cell types. Cells that require endogenous Hippo pathway output, such as squamous cells, have in common being mechanically strained. Thus, Hippo signaling is involved in cellular responses to mechanical strain and in the differentiation and homeostasis of morphologically challenged cells. These findings correct a long-standing misconception about the role of Hippo signaling in organ growth and reveal the need to re-evaluate our understanding of its function in other contexts, such as in cancer and regeneration. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: georg.halder@vib.be
†These authors contributed equally to this work.
Cite this article as W. Kowalczyk *et al.*, *Science* **378**, eabg3679 (2022). DOI: 10.1126/science.abg3679

READ THE FULL ARTICLE AT https://doi.org/10.1126/science.abg3679

RESEARCH ARTICLE

DEVELOPMENTAL BIOLOGY

Hippo signaling instructs ectopic but not normal organ growth

W. Kowalczyk[1]†, L. Romanelli[1]†, M. Atkins[1,2], H. Hillen[1], C. Bravo González-Blas[3], J. Jacobs[3], J. Xie[1], S. Soheily[1], E. Verboven[1], I. M. Moya[1,4], S. Verhulst[5], M. de Waegeneer[3], L. Sansores-Garcia[1], L. van Huffel[1], R. L. Johnson[6], L. A. van Grunsven[5], S. Aerts[3], G. Halder[1]*

The Hippo signaling pathway is widely considered a master regulator of organ growth because of the prominent overgrowth phenotypes caused by experimental manipulation of its activity. Contrary to this model, we show here that removing Hippo transcriptional output did not impair the ability of the mouse liver and *Drosophila* eyes to grow to their normal size. Moreover, the transcriptional activity of the Hippo pathway effectors Yap/Taz/Yki did not correlate with cell proliferation, and hyperactivation of these effectors induced gene expression programs that did not recapitulate normal development. Concordantly, a functional screen in *Drosophila* identified several Hippo pathway target genes that were required for ectopic overgrowth but not normal growth. Thus, Hippo signaling does not instruct normal growth, and the Hippo-induced overgrowth phenotypes are caused by the activation of abnormal genetic programs.

The Hippo pathway is a conserved signal transduction pathway that regulates gene expression through its downstream effectors, the transcriptional coactivators Yap and Taz [Yorkie (Yki) in *Drosophila*] that bind to TEAD [Scalloped (Sd) in *Drosophila*] and other transcription factors (*1*–*5*). The core of the pathway comprises a kinase cascade in which the Mst1/2 kinases [Hippo (Hpo) in *Drosophila*] activate the Lats1/2 kinases [Warts (Wts) in *Drosophila*], which then inhibit Yap/Taz/Yki by phosphorylation (fig. S1A). Hippo pathway activity is modulated by diverse biochemical and mechanical cell-cell and extracellular matrix (ECM)–cell signals integrating information about cell shape, cell adhesion, and tissue integrity (*1*–*5*).

The Hippo pathway is thought to be a master regulator of organ growth (*5*–*10*). This model is largely based on the observation that mutations in Hippo pathway components can lead to organ overgrowth in flies and mice. In particular, the mouse liver and *Drosophila* imaginal disc–derived structures such as eyes and wings show substantial overgrowth upon Hippo pathway deregulation. There, loss-of-function mutations in the core kinases or overexpression of Yap/Taz/Yki up-regulates the expression of genes that induce cell proliferation, cell survival, and tissue growth (*6*, *11*–*20*). Accordingly, the central hypothesis of the Hippo pathway growth control model is that Yap/Taz/Yki are active and drive cell proliferation during the growth phase of an organ, but are then inactivated when the organ has reached its proper size. Thus, in this model, the Hippo pathway acts instructively in that it directs cells when to proliferate and when to stop proliferation.

This model makes three key predictions: (i) Hippo pathway output is required for normal growth; (ii) Hippo pathway activity is modulated during development and correlates with cell proliferation; and (iii) the genetic program that drives overgrowth in Hippo mutants resembles that of normal growth. Consistent with these predictions, loss of Yki results in smaller eyes and wings (*21*), and hyperactivation of Yap/Taz/Yki can drive excessive cell proliferation (*6*, *12*–*20*). However, some findings do not fit the model, namely that *Yap/Taz* mutant hepatocytes and *sd* mutant eye disc cells can proliferate (*22*–*24*). We therefore re-evaluated the function of Hippo signaling during organ growth. We studied *Drosophila* imaginal discs and the mouse liver, because these are major developmental paradigms in which the Hippo pathway and organ growth have been investigated (*10*, *25*, *26*).

Yap/Taz are not required in liver progenitor cells for liver growth

We started by analyzing the phenotype of mouse livers in which we removed Hippo pathway transcriptional output. We conditionally deleted *Yap* and *Taz* during embryonic liver development using floxed alleles (*27*, *28*). We and others have previously reported that codeletion of *Yap* and *Taz* from hepatoblasts, the embryonic liver progenitor cells that give rise to hepatocytes and cholangiocytes (also known as biliary epithelial cells), using *Albumin-Cre (Alb-Cre)* did not affect liver size (*23*, *24*). However, because *Alb-Cre* triggers recombination only after embryonic day 15.5 (E15.5), when hepatoblasts are already proliferating (*29*), we used *Alfp-Cre*, which is expressed earlier, before day E10.5 (*30*), at the onset of hepatoblast specification and before hepatoblasts undergo extensive proliferation (*31*). *Alfp-Cre* efficiently recombined a *R26-Lox-STOP-Lox-tdTomato* reporter transgene starting at E10.5 (fig. S1B) and knocked out *Yap/Taz*, as verified by immunohistochemistry (fig. S1D). *Alfp-Cre;Yap/Taz* double knockout mice (*Alfp-Cre;Yap/Taz*^{KO}) were born at Mendelian ratios but presented yellow ears and paws and had high aspartate transaminase (AST), alanine aminotransferase (ALT), and total bilirubin serum levels, indicating liver damage (fig. S1E). However, gross liver morphology, liver to body weight ratio, and hepatocyte size and numbers were normal (Fig. 1, A to C, and fig. S1C). Rather, mutant mice lacked bile ducts similar to but stronger than the phenotype of *Yap* (and *Taz*) knockout by *Alb-Cre* (*23*, *24*, *32*). This was because few cholangiocytes were specified in *Alfp-Cre;Yap/Taz*^{KO} embryos (Fig. 1D and fig. S1D). Therefore, *Yap* and *Taz* are not required in hepatoblasts for their proliferation but are required for their differentiation into cholangiocytes.

Drosophila eye discs can grow without Hippo pathway transcriptional output

Unlike the ability of the mouse liver to grow in the absence of Yap and Taz, deletion of *yki* in *Drosophila* imaginal disc cells results in smaller eyes and wings (*21*). However, although this phenotype shows that Yki is required for imaginal disc development, such small tissue phenotypes result not only from mutations in genes that control growth, but also from mutations in genes required for basic cellular functions (i.e., housekeeping genes). Therefore, the *yki* loss-of-function phenotype does not reveal whether Hippo signaling acts instructively to command when and where cells proliferate or permissively to simply enable cells to function properly. We therefore turned to Sd, the main transcription factor binding to Yki. Sd is required for hyperactivated Yki to cause overgrowth (*22*, *33*, *34*), but unlike *yki* mutant cells, *sd*-null mutant cells survive and can proliferate (*22*, *33*) (Fig. 1E). As previously reported, this difference in phenotype is explained because Sd acts as a transcriptional repressor in the absence of Yki (*35*). Thus, *yki* mutant cells die because Sd represses cell survival genes in the absence of Yki, but *sd* single or *yki;sd* double mutant cells are able to proliferate because the repressive activity of Sd is removed

[1]VIB Center for Cancer Biology and KU Leuven Department of Oncology, KU Leuven, Leuven, Belgium. [2]Department of Biological Sciences, Sam Houston State University, Huntsville, TX, USA. [3]VIB Center for Brain and Disease Research and KU Leuven Department of Neurosciences, KU Leuven, Leuven, Belgium. [4]Facultad de Ingeniería y Ciencias Aplicadas, Universidad de Las Américas, Quito, Ecuador. [5]Department for Cell Biology, Faculty of Medicine and Pharmacy, Vrije Universiteit Brussel, Brussel-Jette, Belgium. [6]The University of Texas MD Anderson Cancer Center, Houston, TX, USA.
***Corresponding author. Email: georg.halder@vib.be**
†These authors contributed equally to this work.

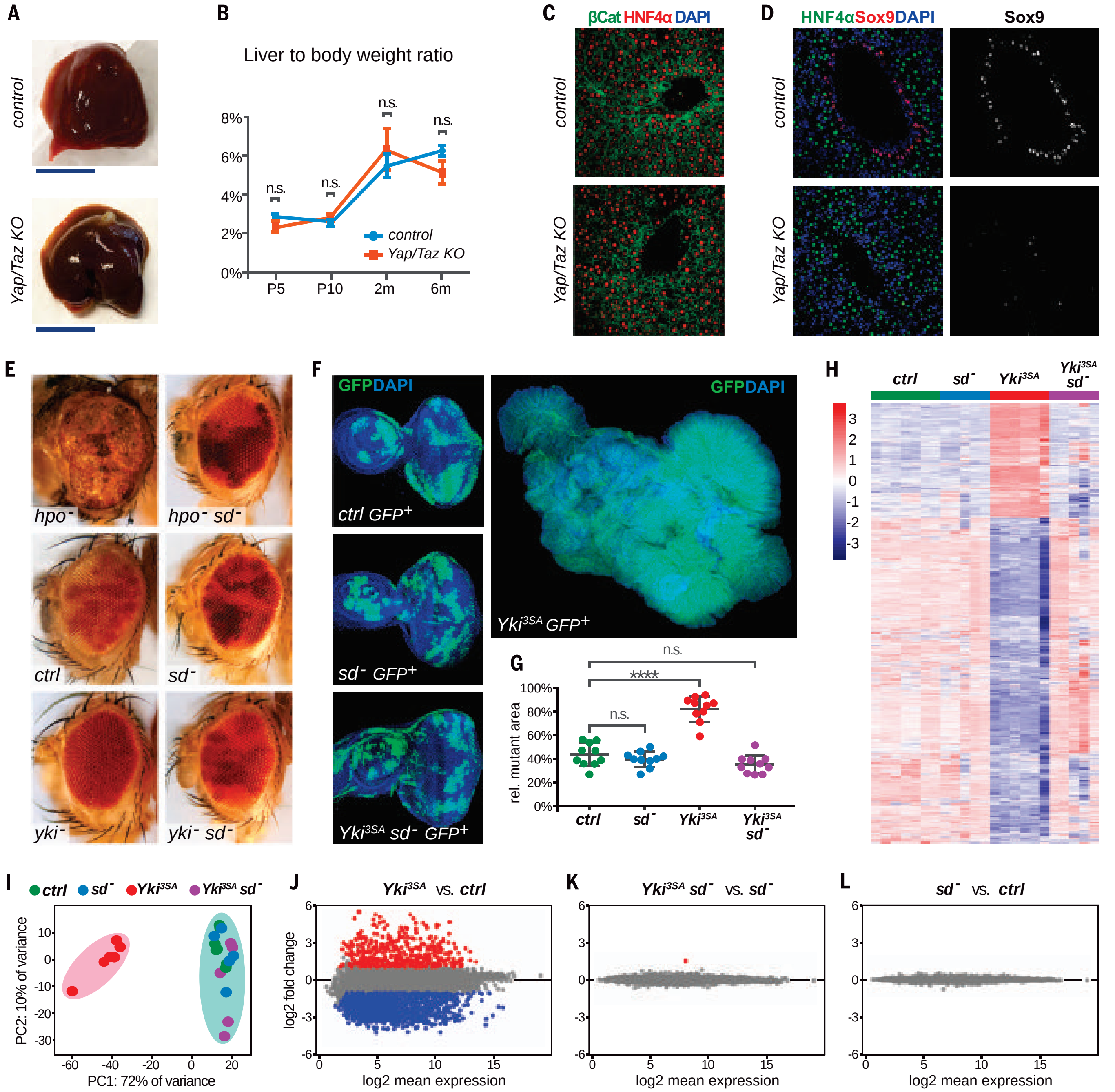


Fig. 1. Loss of Hippo signaling output does not affect developmental growth. (**A**) Control and *Aflp-Cre;Yap/Taz KO* whole liver pictures from adult mice (6 months old; scale bars, 1 cm). (**B**) Average liver to body weight ratios from postnatal day 5 to adulthood (N = 3). (**C** and **D**) Liver sections of indicated genotypes stained for nuclei [4′,6-diamidino-2-phenylindole (DAPI), blue], adherens junctions [β-catenin, green in (C)], hepatocytes [HNF4α, red in (C), green in (D)] and cholangiocytes [Sox9, red in (D)]. (**E**) Lateral views of mosaic adult *Drosophila* eyes with mutant clones of the indicated genotypes; mutant cells are marked by the absence of red pigment. (**F** and **G**) Immunofluorescent images and quantification of clone size of eye imaginal discs with clones of cells of the indicated genotypes marked by GFP expression. (**H**) Heatmap showing relative expression levels of genes differentially expressed between Yki^{3SA} and control eye discs (adjusted $P < 0.05$; $|log2FC|>1$) of the indicated genotypes. (**I**) Principal components analysis plot showing all samples of indicated genotypes. (**J** to **L**) MA plots for the indicated comparisons; up-regulated genes (adjusted $P < 0.05$; log2FC > 1) are shown in red, and down-regulated genes (adjusted $P < 0.05$; log2FC < –1) are shown in blue. ****$P < 0.0001$; n.s., not significant. Error bars indicate SD.

(Fig. 1E, bottom). Therefore, *sd* mutants provide a tool to eliminate Hippo pathway output without killing the cells. However, because Yki can also interact with other transcription factors (*36*, *37*), we first determined to what extent deletion of Sd reduces the ability of Yki to drive gene expression and thus Hippo pathway transcriptional output.

To this end, we studied eye imaginal discs rather than wing discs, because Sd controls the fate of wing disc cells in addition to its role in the Hippo pathway (*38*). We induced mutant clones early in eye development, before the major growth phase, using the *eyeless-Flipase (ey-Flp)*/FRT MARCM recombination

system, and inspected phenotypes in adult flies or third instar discs, in which most of the eye disc cell proliferation has already occurred. As reported previously (22, 33, 34), deletion of *sd* suppressed the overgrowth caused by loss of *hpo* (Fig. 1E) and the immense overgrowth of cell clones that overexpressed Yki^{3SA}, a constitutively active form of Yki that has all three Warts phosphorylation sites mutated to alanine (39) (Fig. 1, F and G). At the level of the transcriptome, deleting *sd* completely inhibited the ability of Yki^{3SA} overexpression to cause changes in gene expression (Fig. 1, H to K; fig. S2, A to C; and table S1). Thus, although Yki^{3SA} overexpression caused differential expression of 1632 genes in wild-type third-instar eye discs [log2-fold change (log2FC) > 1; adjusted $P < 0.05$] (Fig. 1J and fig. S2B), these effects were fully suppressed when Yki^{3SA} was overexpressed in *sd* mutant clones (Fig. 1K and fig. S2, A to C). Such double mutant discs differentially expressed only the *rosy* gene, which is an eye color gene that does not affect growth. Therefore, Yki hyperactivation has no effect on gene expression and cell proliferation anymore when Sd is not present. Therefore, these results show that deletion of *sd* abolishes the transcriptional output of the Hippo pathway and provides a model with which to study imaginal disc growth control in the absence of Hippo output.

We then analyzed *sd* mutant phenotypes in more detail in third-instar eye discs because they allow visualization of cell proliferation, tissue patterning, and cellular differentiation. Third-instar eye discs are flat, sac-like epithelial structures in which one side, the disc proper, is a columnar epithelium (of ~40,000 cells) that gives rise to the adult eye, antenna, and head cuticle, whereas the other side, the peripodial epithelium, comprises large squamous cells (~2000) that are required for disc eversion during metamorphosis (Fig. 2A) (40–42). *sd*-null mutant clones in the disc proper had normal numbers and distribution of cells in the S phase and grew to normal size and without developmental delay (Fig. 2, B to D). They also had normal patterning and neuronal specification (Fig. 2D). Similarly, knockdown of *sd* specifically in the growing retinal field by expressing upstream activator sequence RNA interference (UAS-RNAi) transgenes using the *optix-Gal4* driver did not reduce eye size or affect morphology (Fig. 2E and fig. S3A), nor did it increase the variation in eye size between the left and right eyes of individual flies, a measure of fluctuating asymmetry that evaluates the robustness of growth control (43) (Fig. 2F). *optix>sd*^{RNAi} knockdown was effective because coexpressing *sd*-RNAi reversed the small eye phenotype of *yki* knockdown (Fig. 2E). Thus, loss of *sd* did not detectably affect the development of the disc proper. By contrast, *sd* mutant clones were not recovered in the peripodial epithelium (Fig. 2, B and C), and knockdown of *sd* specifically in peripodial cells with *C311-Gal4* produced eye discs without peripodial epithelium but with folded retinal fields and pharate adults with severe defects in head eversion (40), typical for defects in peripodium function (Fig. 2, G to I). The lack of peripodial cells in *C311>sd*^{RNAi} eye discs was caused by defects in cell proliferation and survival rather than a transformation into retinal cells, because coexpression of the antiapoptotic protein p35 partially rescued the loss of the peripodial cells, which, however, produced a clump of morphologically abnormal cells (fig. S3B).

At the transcriptional level, the expression profile of eye discs with *sd* mutant clones was not significantly different from that of control discs (Fig. 1L). Thus, loss of Sd did not affect the expression of cell cycle genes or the basal expression levels of Hippo target genes. Correspondingly, loss of Sd reverted the massive increase in Yki target gene expression caused by Yki^{3SA} overexpression back to wild-type levels in *sd;Yki*^{3SA} double mutants (fig. S2, A to C). We did not observe effects on peripodial gene expression because these discs did not have *sd* clones in the peripodial epithelium (Fig. 2, B and C, and materials and methods). Altogether, Sd, and thus transcriptional output from the Hippo pathway, is not measurably required for the growth of the eye disc proper but is required for the development of the peripodial epithelium.

Yap/Taz activity during liver development

The second prediction of the model was that Hippo activity is modulated during development such that Yap/Taz/Yki activity correlates with cell proliferation. To test this, we measured the activity of Yap/Taz by analyzing the expression pattern of a stringent Yap/Taz target gene signature (for simplicity "Yap signature") of 22 genes (44). The Yap signature was strongly up-regulated in response to Yap hyperactivation in adult hepatocytes, confirming that it reports Yap activity (fig. S4, A to E). In addition, we mapped the signature onto an extensive single-cell RNA-sequencing (RNA-seq) atlas of the adult mouse liver that included parenchymal cells, endothelial cells, fibroblasts, immune cells, and other cell types (45). The Yap signature was preferentially expressed in cholangiocytes, endothelial cells, and fibroblasts, which are precisely the cell types in the liver that require Yap/Taz for their development (fig. S4, F and G) (46). We then mapped the Yap signature and a canonical cell cycle signature onto a single-cell RNA-seq dataset of embryonic liver development that encompasses the stages when hepatoblasts proliferate and start to differentiate into hepatocytes and cholangiocytes (E10.5 to E17.5) (Fig. 3, A and B) (47). As expected, expression of the cell cycle genes was strongly enriched in early (proliferating) hepatoblasts and in a small cluster of proliferating cholangiocytes (Fig. 3C and fig. S5A). By contrast, the Yap signature was not enriched in hepatoblasts but was enriched in all cholangiocytes (Fig. 3D and fig. S5, B to D). Yap activity was comparable in cycling and noncycling cholangiocytes, and a single-cell scatter plot further showed a general lack of correlation between Yap signature expression and cell proliferation (Fig. 3E and fig. S5C). The pattern of Yap signature expression was highly robust and driven by most of the Yap signature genes (fig. S5, B, D, and E). Therefore, the pattern of Yap activity during liver development mirrors the phenotype of *Yap/Taz* mutants in cholangiocyte specification and cements that Yap/Taz drive cholangiocyte development but do not control hepatoblast proliferation and thus liver growth.

Yki activity during eye disc development

We also compared the pattern of Yki activity and cell proliferation during *Drosophila* imaginal disc development. To measure Yki activity, we generated a Yki target gene signature by assembling a list of high-quality Sd-binding regions in the genome by overlapping Sd chromatin immunoprecipitation-sequencing (ChIP-seq) peaks and Sd-binding motifs] and then selecting nearby genes that were up-regulated by Yki activation in eye and wing discs (fig. S6, A and B, and materials and methods). The resulting gene signature contained 52 genes, including the Yki target genes *ex*, *kibra*, and *wts* (table S2), and the signature's expression level correlated with the strength of Yki activity in different Hippo pathway mutants (fig. S7, A and B). Therefore, the Yki signature reports Yki activity in imaginal discs.

We then analyzed the expression of the Yki signature using an RNA-seq dataset of discs at different time points during development spanning from mid-second-instar [72 hours after egg laying (AEL)] to late-third-instar (120 hours AEL) just before puparium formation. The expression of >40% of the Yki signature genes increased at later time points, but, overall, the expression patterns were not uniform and poorly correlated with time (fig. S7C and table S3), suggesting that Yki activity is not significantly down-regulated toward the end of disc growth. However, a potential down-regulation of Yki activity may have been obscured, because cells in the posterior eye disc had already started differentiating while more anterior cells were still proliferating (Fig. 2D, *ctrl*, and fig. S7D), thereby never providing a stage at which all cells synchronously ceased proliferation. To circumvent this problem, we fed larvae a special diet that lacks precursor molecules necessary for the synthesis of the molting hormone ecdysone, which is required for pupation, thus keeping animals in the larval stage (48). The

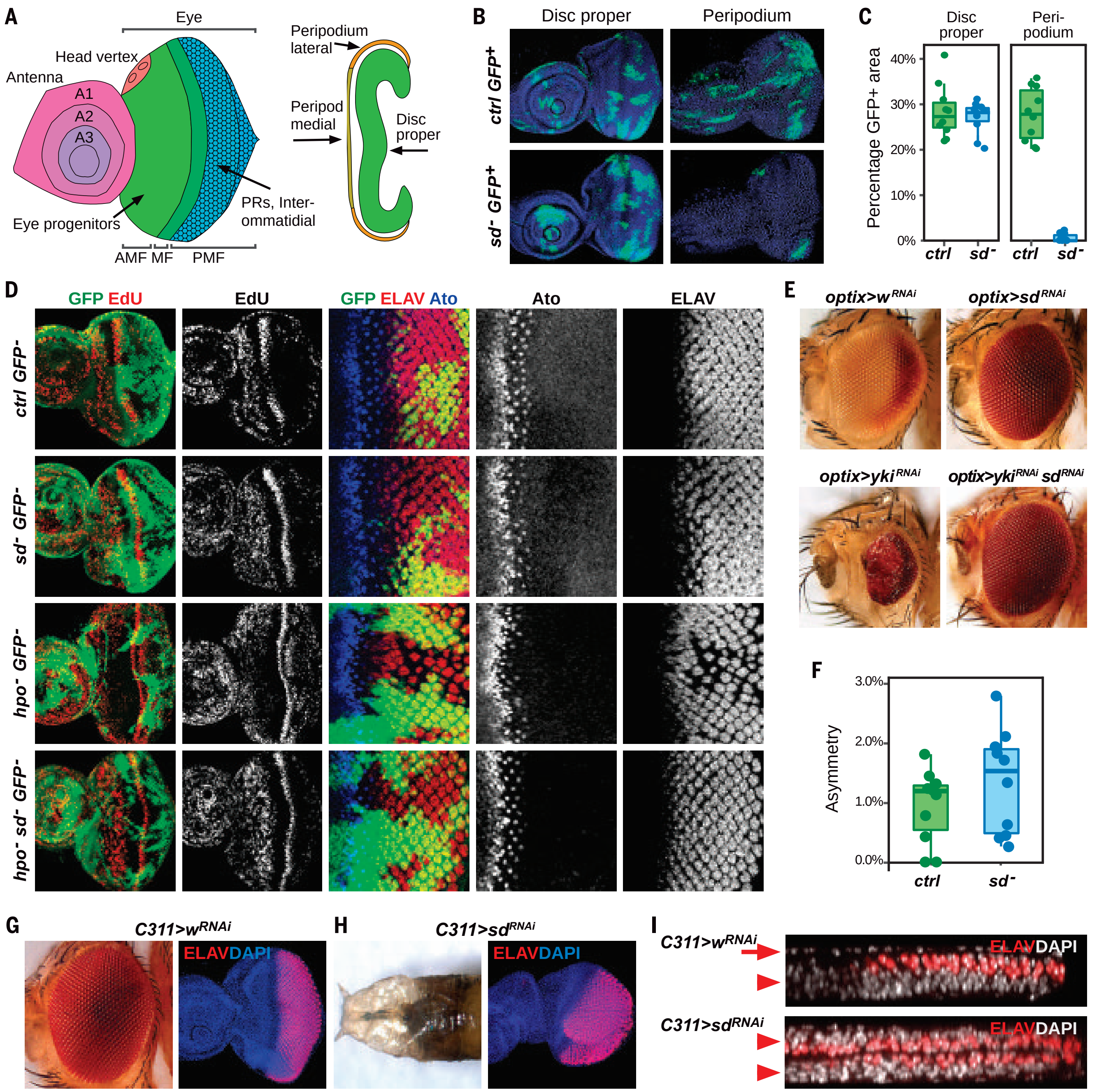


Fig. 2. Sd is required for the development of the peripodial epithelium but not for the disc proper. (**A**) Top view and cross section schematic representation of a wild-type third-instar eye antennal disc showing the location of its major cell types. PR, photoreceptor; MF, morphogenetic furrow; AMF, anterior to the MF; PMF, posterior to the MF. (**B** and **C**) Representative images and size quantification of control and *sd*⁻ mutant clones in disc proper and peripodium. Clones are marked by GFP (green) (N = 10). (**D**) Eye imaginal discs with mutant clones of the indicated genotypes marked by the absence of GFP (green) expression. EdU (red) incorporation labels cells in the S phase. Details of eye imaginal discs with anti-Atonal (Ato; blue) and anti-ELAV (red) staining illustrates neuronal patterning and photoreceptor differentiation. (**E**) Lateral views of adult *Drosophila* eyes of the indicated genotypes. (**F**) Quantification of asymmetry (difference in the number of ommatidia between the two eyes of the same animal) in control and *sd*⁻ mutant retinae (N = 10, P = 0.25). (**G** and **H**) Lateral views of an adult head and a pupa lacking a head and confocal images of eye imaginal discs of the indicated genotypes stained with DAPI (blue) and anti-ELAV (red) to label nuclei and photoreceptors, respectively. (**I**) Cross-section view of the eye imaginal discs shown in (G) and (H). Arrow and arrowheads point to the peripodium and disc proper, respectively.

imaginal discs in these larvae grew until their terminal size was reached but did not proceed with differentiation because this requires a pulse of ecdysone. This method thus allowed us to directly compare actively growing discs with discs at the same developmental stage that terminated proliferation after reaching their proper size. Indeed, 5-ethynyl-2′-deoxyuridine (EdU) incorporation was nearly absent, and RNA-seq showed that cell cycle genes were down-regulated in growth-completed discs from 11-day-old nonpupating larvae (Fig. 3, F and G, and fig. S8, A to D). However, of the Yki signature genes, only eight (15%) were down-regulated and >40% (23) were actually

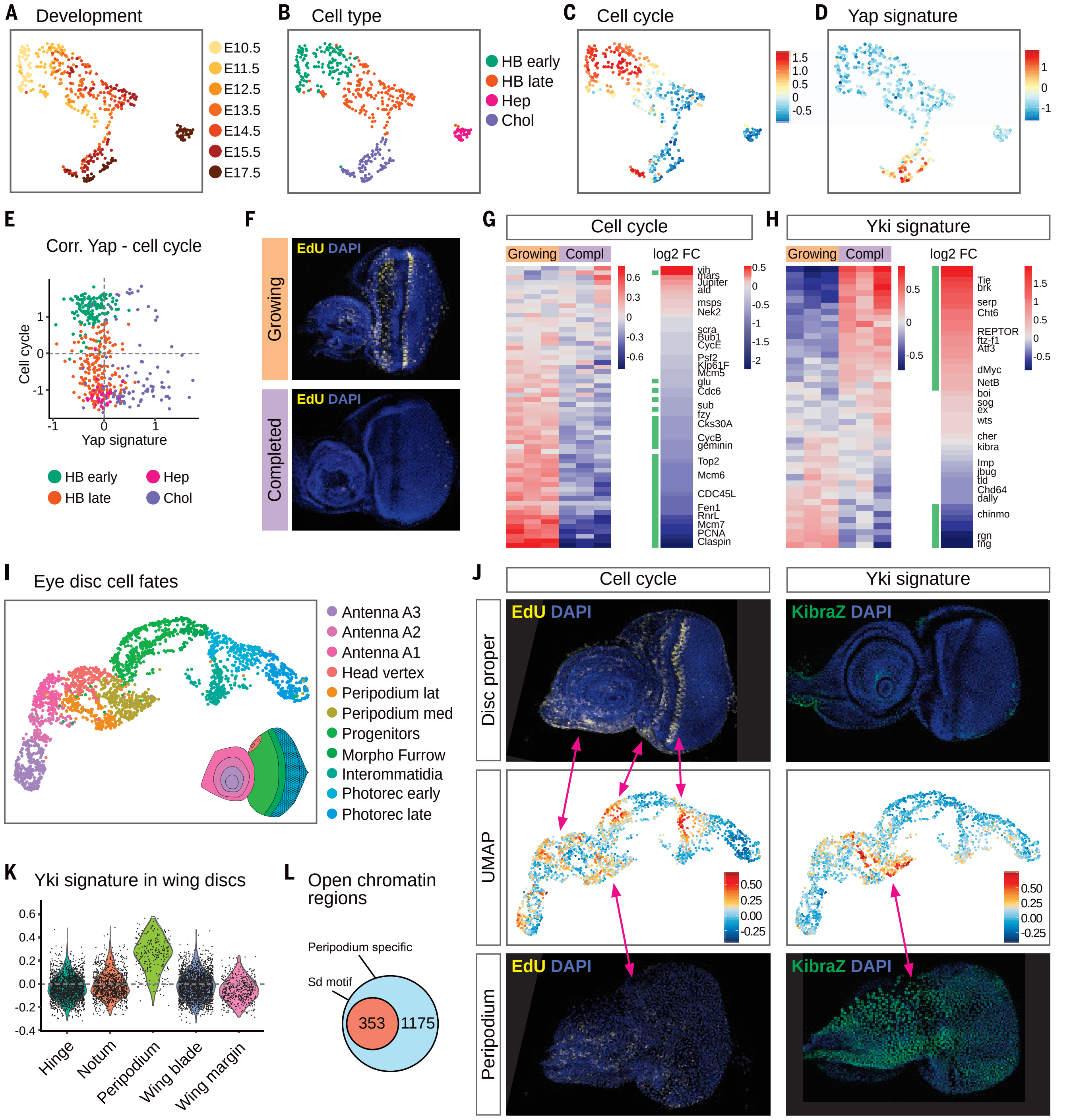


Fig. 3. Yap/Taz/Yki activity does not correlate with cell proliferation. (**A** and **B**) Uniform manifold approximation and projection (UMAP) plot of single-cell RNA-seq data from embryonic murine livers colored by developmental time point (A) and cell type (B). HB, hepatoblast; Hep, hepatocyte; Chol, cholangiocyte. (**C** and **D**) Average expression of cell cycle genes (C) and Yap signature genes (D). (**E**) Scatter plot showing the correlation between Yap signature and cell cycle gene expression in single cells colored by cell identity. (**F**) Actively growing (day 5 AEL) and growth-completed (day 11 AEL) eye imaginal discs from animals feed a pupation-preventing diet, stained for nuclei (DAPI, blue) and EdU (yellow) marking cells in S-phase. (**G** and **H**) Heatmaps showing relative expression and log2FCs (green sidebar marks statistically significant changes; adjusted $P < 0.05$) in cell cycle genes (G) and Yki signature genes (H) between actively growing and growth completed eye imaginal discs. (**I**) UMAP plot of single-cell RNA-seq data with 2596 cells colored by assigned cell type together with a schematic visualization of an eye imaginal disc. lat, lateral; med, medial; morpho, morphogenetic; photorec, photoreceptor. (**J**) Left column: cell proliferation in an eye disc shown by EdU incorporation (yellow) in the disc proper (top) and peripodium (bottom) and average expression of cell cycle genes on the UMAP (middle). Right column: Yki activity shown by kibra-lacZ staining (green) in the disc proper (top) and peripodium (bottom) and average expression of Yki signature genes on the UMAP (middle). The arrows point to corresponding groups of cells. (**K**) Violin plot showing average expression of Yki signature genes in single cells of the wing imaginal disc grouped by cell type. (**L**) Venn diagram showing the proportion of all chromatin regions specifically accessible in peripodial cells overlapping with those that also show enrichment for Sd-binding motifs.

up-regulated in growth-completed discs (Fig. 3H; fig. S8, E and F; and table S3). Moreover, knockdown of *yki* reduced the expression of the Yki reporter gene *ex-lacZ* in growing as well as growth-completed discs (fig. S9). These results show that Yki is still active in growth-completed discs and not inactivated when growth ceases. This was also observed using a similar approach in wing discs (*49*).

In a second approach, we separated cell proliferation from differentiation by using a single-cell gene expression atlas of third instar eye discs (*50*). This atlas reproduces the switch from proliferating precursor cells in the anterior to differentiating photoreceptor and interommatidial support cells in the posterior of the eye disc (Fig. 3I). The cell cycle signature was preferentially expressed in proliferating antennal and retinal precursor cells and in cells in the second mitotic wave, which is consistent with the proliferation pattern observed by EdU incorporation (Fig. 3J, left, and fig. S10A). By contrast, expression of the Yki signature was highest in peripodial cells and showed no correlation with cell proliferation (Fig. 3J, right, and fig. S10, B to E). Likewise, the Yki signature was also preferentially expressed in peripodial cells of wing discs (Fig. 3K and fig. S11, A to F).

To further scrutinize our genomics findings, we examined Yki activity in situ. Yki intracellular localization, assayed by a functional Yki–green fluorescent protein (Yki-GFP) genomic knock-in construct, showed a tendency for enhanced nuclear localization in peripodial cells (fig. S11H), but this assay was not very sensitive (*51*). Conversely, the expression of *ex-lacZ* was high in peripodial cells, but it showed Sd independent basal expression in disc proper cells (*22*), which prevented its use to compare Yki activity in peripodial and disc proper cells (fig. S11I). As previously observed and shown in our RNA-seq data, *kibra* expression strongly correlated with Yki activity (fig. S11J) (*52*). We thus monitored *kibra* expression with a *lacZ* enhancer trap insertion into the *kibra* gene. *kibra-lacZ* expression was strongly induced upon Yki activation by *wts* knockdown (fig. S11, K and L), indicating that it is regulated by Yki. In wild-type animals, *kibra-lacZ* expression was absent in disc proper cells but high in peripodial cells, even when cell proliferation stopped after disc growth was complete (Fig. 3J, right, and fig. S11M). Thus, *kibra-lacZ* is a new Yki-Sd activity reporter that reports positive Yki-Sd output, and its expression pattern confirmed the predominantly peripodial activity of Yki-Sd.

The preferential activity of Yki-Sd in peripodial cells was also observed at the chromatin level. Analysis of our single-cell assay for transposase-accessible chromatin sequencing (ATAC-seq) dataset of accessible chromatin in eye disc cells (*50*) showed that >90% of the Sd-binding regions were accessible in peripodial cells, and 25% were exclusively accessible in peripodial cells (fig. S11G). In addition, >20% of all peripodial-specific accessible chromatin regions were bound by Sd (Fig. 3L). Therefore, peripodial specific enhancers are strongly enriched for Sd binding, and Sd-bound enhancers are preferentially accessible in peripodial cells. Altogether, the data presented in this section show that the activity of Yki-Sd does not correlate with the rate of cell proliferation, but does correlate with the development of peripodial cells, which is consistent with the loss-of-function phenotype of *sd* mutants.

Yap hyperactivation induces an ectopic program in hepatocytes

The third prediction of the model positing that the Hippo pathway is a master regulator of organ growth is that the overgrowth induced by Yap/Taz/Yki hyperactivation recapitulates growth during normal development. To test this prediction, we compared the genetic program active during Yap/Taz/Yki–driven overgrowth with that of normal development. We studied how ectopic Yap activation causes liver overgrowth in mice. We conditionally activated Yap in hepatocytes of adult mice using two different methods to ensure the highest robustness of our results. We deleted the *Lats1/2* kinases by injecting AAV8-Cre into *Lats1/2* double-floxed mice (*Lats1/2*^{KO}) and overexpressed YAP by feeding doxycycline to mice that expressed the tetracycline-activated transactivator (rtTA) under the hepatocyte-specific *ApoE* promoter and human YAP from a rtTA-responsive TetO promoter (hereafter *Apo>YAP*) (*6*). Adult hepatocytes are normally quiescent, but Yap activation triggered cell cycle re-entry and caused the liver to more than double its size within 2 weeks in both mutants, as previously described (fig. S12A) (*6*). One week after Yap activation, purified *Lats1/2*^{KO} and *Apo>YAP* hepatocytes had massive changes in gene expression that were highly correlated (R = 0.84; Fig. 4A and fig. S12, B to F). The overlap comprised 1819 up-regulated and 372 down-regulated genes (log2FC > 1, adjusted P < 0.05; table S4). Functions of up-regulated genes (hereafter called the Yap^{hyper} program) were enriched for cell cycle regulation (113 genes, 6%), ECM regulation, cell motility and angiogenesis (306 genes, 17%), and immune functions (417 genes, 23%), (fig. S12, G and H), whereas the down-regulated genes were enriched for hepatocyte metabolic functions (fig. S12, I and J). The Yap^{hyper} program is different from the Yap signature in that it included all 1819 genes that were either directly or indirectly up-regulated in hepatocytes in response to Yap hyperactivation, whereas the Yap signature is a very select set of 22 genes that are broadly activated in response to Yap.

We then investigated to what extent the Yap^{hyper} program resembled that of embryonic liver development. We mapped the Yap^{hyper} program onto the single-cell RNA-seq atlas of liver development (*47*). The average expression of Yap^{hyper} program genes was highest in proliferating cholangiocytes, followed by noncycling cholangiocytes and early (proliferating) hepatoblasts (Fig. 4B). However, this pattern was a composite of the expression patterns of four gene clusters with distinct expression patterns (Fig. 4C). The first cluster of 247 genes (13% of the Yap^{hyper} program) contained genes that were expressed in all proliferating cells, mostly early hepatoblasts and a few proliferating cholangiocytes. The second cluster of 312 genes (17% of the Yap^{hyper} program) was expressed specifically in all cholangiocytes regardless of the cell cycle. The other two clusters had uniform or overall low expression in embryonic liver cells. Thus, although the expression of Yap^{hyper} genes in hepatoblasts showed that Yap induces parts of the hepatoblast program, this was not because Yap induced progenitor cell characteristics but rather because it induced general cell cycle genes. Indeed, the Yap^{hyper} program contained <5% of the hepatoblast specific gene set (fig. S12K). Conversely, Yap activated genes that are specifically expressed in embryonic cholangiocytes. However, the Yap^{hyper} program represented only ~15% of all cholangiocyte markers (fig. S12K). It follows that Yap hyperactivation in adult hepatocytes did not simply activate the genetic program of embryonic hepatoblasts but did induce a general cell proliferation program and partially transdifferentiated hepatocytes toward cholangiocyte fate.

Another possibility was that Yap hyperactivation induced a program that is used during liver regeneration, when hepatocytes re-enter the cell cycle. To test this, we performed RNA-seq on regenerating hepatocytes that we purified 48 hours after acute injury by injection of the liver toxin CCl_4, when hepatocyte proliferation peaks. Regenerating hepatocytes up-regulated 1021 genes, of which more than half were also present in the Yap^{hyper} program (fig. S14A). Thus, Yap hyperactivation indeed induced a large portion of the liver regeneration program (Fig. 4, D and G, and fig. S13A). However, endogenous Yap/Taz were not required for hepatocyte proliferation (*23*) or the induction of the regeneration program after CCl_4 injection (Fig. 4, E and F; fig. S13, B to E; and table S5). This conundrum can be resolved because Yap hyperactivation and the resulting overgrowth caused liver damage, as revealed by elevated serum levels of AST (fig. S13G), which would then induce the regeneration program as a secondary effect. Therefore, the Yap^{hyper} program comprises direct effects caused by Yap hyperactivation and indirect effects triggered by liver damage.

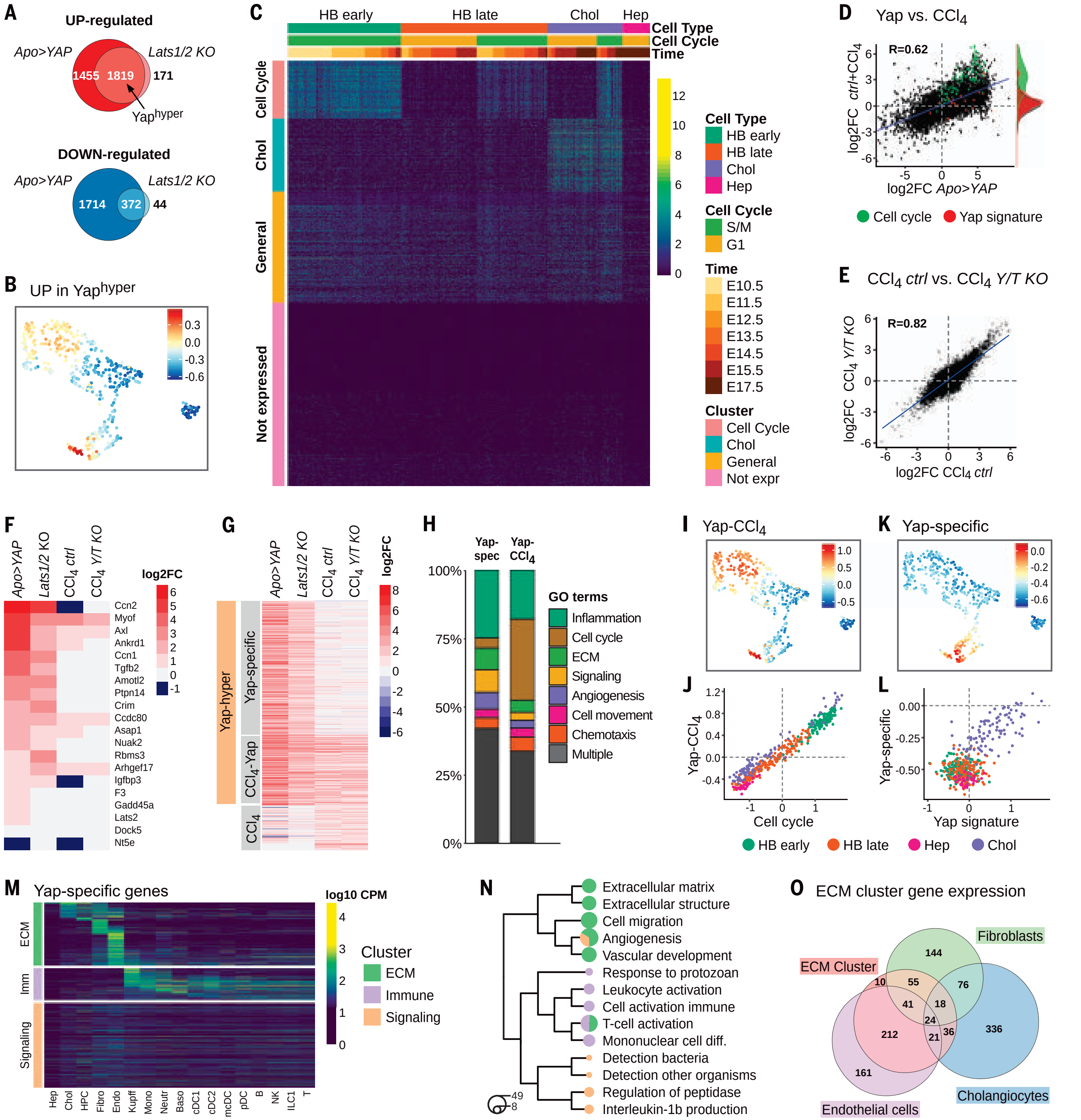


Fig. 4. Yap hyperactivation drives an ectopic growth program in hepatocytes. (**A**) Overlap between the up-regulated (top) and down-regulated (bottom) genes in hepatocytes with YAP overexpression and *Lats1/2 KO*. (**B**) Average expression of Yap^{hyper} program genes in the developing liver. (**C**) Clustering of Yap^{hyper} program genes by their normal expression pattern during liver development. Cells (columns) are grouped by cell type, cell cycle phase, and developmental time; genes (rows) are grouped by similar expression pattern. (**D**) Correlation of changes in gene expression between *Apo>YAP* and CCl_4 (compared with their relative controls). Cell cycle and Yap signature genes are marked in green and red, respectively, and their distribution is shown by the density plot on the right side. (**E**) Correlation of changes in hepatocyte gene expression induced by CCl_4 injection into control (*ctrl*) and *Yap/Taz KO* animals. (**F**) log2FCs of Yap signature genes for comparisons between the indicated conditions and their respective controls; values are rounded to the nearest integer. (**G**) log2FC of genes (rows) induced by Yap hyperactivation or CCl_4 injection for comparisons (columns) between the indicated conditions and their respective controls. (**H**) Barplot showing the percentage of genes in the Yap-specific or Yap-CCl_4 subsets of the Yap^{hyper} program that belong to the given Gene Ontology (GO) term categories. (**I** to **L**) Average expression of the Yap-CCl_4 (I) and Yap-specific (K) parts of the Yap^{hyper} program in single cells of the embryonic liver and its correlation with the expression of cell cycle (J) and Yap signature (L) genes, respectively. (**M** and **N**) Subclustering of the Yap-specific genes by their expression in different cell types of the adult liver. (M) Heatmap showing the absolute expression [log10 counts per million (CPM)] in single cells aggregated by cell type. (N) Treeplot showing the top five GO terms enriched for each cluster. Dot size corresponds to the number of genes, and dot color reflects the proportion of genes in each cluster. (**O**) Overlap between genes in the ECM cluster and marker genes of fibroblasts, cholangiocytes, and endothelial cells.

Therefore, we divided it into a general regeneration part that comprised the genes induced by both Yap and CCl_4 (Yap-CCl_4) and a "Yap-specific" part that comprised the genes that were induced by Yap but not by CCl_4 (Fig. 4G and table S6). Genes in the Yap-CCl_4 set predominantly function in the cell cycle and immune system, are the most active in early hepatoblasts and proliferating cholangiocytes, and their expression is poorly correlated with endogenous Yap activity (Fig. 4, H to J). This set thus constitutes the regenerative response to the liver damage caused by Yap-induced liver overgrowth. By contrast, genes in the Yap-specific set were normally expressed in embryonic cholangiocytes (Fig. 4, K and L), as well as in endothelial cells, fibroblasts, and cholangiocytes of adult livers (fig. S14). These are the cells with the highest Yap activity and they require Yap/Taz for their development (*46*).

To understand the biological functions of the Yap-specific genes, we classified them according to their expression patterns in normal adult liver. A first cluster of genes (ECM cluster; Fig. 4, M and N) was expressed in endothelial cells, fibroblasts, and cholangiocytes and was strongly enriched for functions in angiogenesis and ECM. A second cluster was predominantly expressed in immune cells and enriched for immune functions (immune cluster; Fig. 4, M and N). A third cluster had uniform but low expression in normal livers (signaling cluster; Fig. 4, M and N). The ECM cluster comprised >50% of the endothelial cell specific markers and 33% and 18% of the fibroblast and cholangiocyte-specific markers, respectively, with several genes expressed in all three cell types (Fig. 4O). Altogether, these data show that Yap hyperactivation in adult hepatocytes does not cause liver overgrowth because it activates a normal progenitor (hepatoblast) cell program. Rather, Yap hyperactivation ectopically induces parts of the genetic programs of endothelial cells, cholangiocytes, and fibroblasts, as well as other genes that are not normally expressed in hepatocytes.

Ectopic Yki activates an abnormal genetic program in imaginal discs

Next, we investigated how ectopic Yki drives tissue overgrowth in *Drosophila* imaginal discs. As we did for the liver, we overexpressed the constitutively active Yki^{3SA} and deleted the *Lats1/2* homolog *wts* in clones of cells. Both conditions produced eye discs that were nearly entirely composed of mutant cells and lacked ommatidial patterning and photoreceptor differentiation (Fig. 1F). Discs had substantial changes in gene expression that were highly correlated between the two mutants (Fig. 5A; fig. S15, A to C; and table S8). Consistent with the histological phenotype, induced genes were mainly involved in ribosome biogenesis (indicating accelerated cell growth) and DNA replication, whereas the down-regulated genes were involved in retinal cell differentiation (fig. S15D). This phenotype raised the possibility that hyperactivation of Yki during eye development trapped cells in an early developmental stage, thereby producing an excess of precursor cells. However, three lines of evidence refute this hypothesis.

First, we compared the Yki^{hyper} program, defined as the 222 genes overlapping between the Yki^{3SA} and wts^- mutant profiles, with that of wild-type eye discs from different developmental stages (*53*). Forty-four percent (n = 98) of the Yki^{hyper} program genes were expressed at higher levels at early stages compared with later stages of eye development (Fig. 5, B and C). However, these genes represented <4% of the genes that characterize early eye development (Fig. 5B and fig. S15E). Therefore, Yki hyperactivation did not simply extend early development.

Second, mapping the Yki^{hyper} program onto the single-cell expression atlas of third-instar eye discs showed that Yki^{hyper} cells down-regulated genes normally expressed in differentiating photoreceptor cells, consistent with their lack of retinal differentiation (Fig. 5D). The up-regulated genes, however, were not enriched for genes expressed in progenitor cells but rather for genes expressed in peripodial cells (Fig. 5, E and F). Expression in peripodial cells was the predominant pattern, with few of the Yki^{hyper} program genes being broadly or lowly expressed in normal eye discs (Fig. 5F). Because *ey-Flp* induces clones not only in the disc proper but also in peripodial cells, this result may simply be the consequence of hyperactivating Yki in peripodial cells causing their overproliferation. However, driving Yki^{3SA} overexpression specifically in disc proper cells by *optix-Gal4* caused essentially the same phenotype and triggered the induction of peripodial cell–specific genes (fig. S15, F and G). Likewise, in wing discs, *wts* mutant clones and Yki overexpression induced a genetic program that was most similar to that of peripodial cells (fig. S16). However, Yki induced less than half of the peripodial cell markers (fig. S15, H and I). Therefore, Yki activation did not simply cause an expansion of retinal progenitor cells and inhibit their differentiation, but ectopically activated a portion of the peripodial cell program.

Third, we compared the Yki^{hyper} program with a gene set representative of normal growth that we generated as the difference in gene expression between actively growing and growth completed eye discs collected from larvae fed the ecdysone precursor–deficient food (fig. S15J). There was minimal overlap between the Yki^{hyper} and normal growth programs and, unlike the Yki^{hyper} program, the normal growth program was not specifically expressed in peripodial cells, but rather was expressed broadly and reflected the pattern of cell proliferation (Fig. 5, G, H, K, and L). Conversely, the Yki^{hyper} program resembled abnormal growth, here defined as genes up-regulated in Ras^{V12G}–driven neoplastic tumors (*Ras^{V12G} $scrib^{RNAi}$* tumors) (*55*). Such tumors have hyperactive Yki (*54*), and their gene expression profile significantly overlapped with the Yki^{hyper} program and was enriched in peripodial cells (Fig. 5, I to K, and fig. S17, A to G). Altogether, ectopic Yki activation did not simply stimulate extra normal growth, but partially reprogrammed disc proper cells toward a peripodial cell phenotype (Fig. 5M and fig. S17H).

Next, we investigated which Yki-Sd target genes were required for Yki-driven overgrowth. To this end, we performed a genetic screen in which we assessed the role of the 322 predicted Sd target genes plus an additional 143 low confidence targets (table S10 and materials and methods). We knocked down each of these genes in eye discs that overexpressed Yki^{1SA} under the control of the *GMR-Gal4* driver (*GMR>Yki^{1SA}*), which causes massively overgrown adult eyes (*22*). We thus crossed *GMR>Yki^{1SA}* flies to a collection of 1673 fly stocks that each expressed an RNAi construct targeting one of the candidate genes under the control of a Gal4 responsive UAS-promoter (on average, 3.6 different RNAi lines per gene). Knockdown of 44 genes suppressed *GMR>Yki^{1SA}* overgrowth, whereas knockdown of 40 enhanced overgrowth, and knockdown of the remaining 381 did not noticeably affect the adult eye phenotype (Fig. 5N and table S10). We then screened all 465 genes for requirement during normal growth using the *optix-Gal4* driver, which is active in disc proper cells during eye development. Here, knockdown of 70 genes reduced eye size, whereas knockdown of eight resulted in larger eyes (Fig. 5N and table S10). Overlapping the results of the two screens showed that 25 (57%) of the 44 Yki^{1SA}-overgrowth suppressors were also required for normal growth, but knockdown of the other 19 (43%) suppressors did not affect normal growth (Fig. 5N). Although RNAi knockdown can be incomplete and may not reveal full loss-of-function phenotypes, our results indicate that the *GMR>Yki^{1SA}* phenotype is preferentially sensitive to knockdown of these genes. Thus, these results further support that Yki-driven overgrowth does not simply recapitulate normal growth.

Among the Yki^{1SA} overgrowth suppressors that were also required for normal growth were RNAi lines targeting Yki as expected; epidermal growth factor receptor (EGFR) and its ligands, Vn and Spi; the cell cycle regulators E2f1, CyclinE, and Stg; the cell growth promoter dMyc; and components of Notch signaling, Dl, Gp150, Uif, and E(spl)malpha (Fig. 5O). Overexpression of these genes can cause ectopic cell proliferation, although overexpression

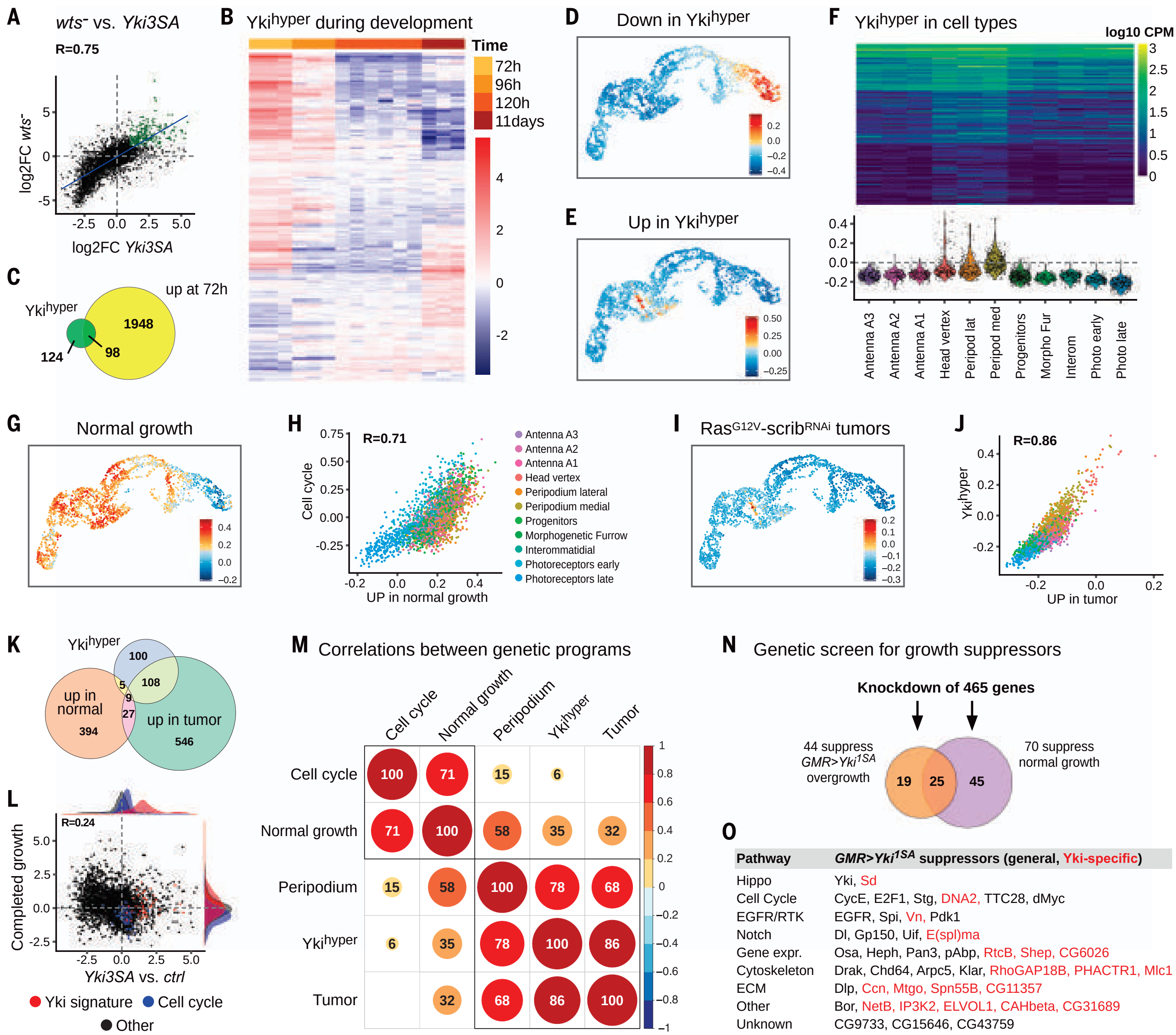


Fig. 5. Yki hyperactivation driven overgrowth resembles tumor but not normal growth. (**A**) Correlation of changes in gene expression between Yki^{3SA} overexpression and *wts KO* (compared with their respective controls). Yki^{hyper} program genes are shown in green. (**B**) Relative expression levels of the Yki^{hyper} program genes at different time points of larval growth (72, 96, and 120 hours AEL) and after growth completion (11 days AEL). (**C**) Overlap between the Yki^{hyper} program and the genes up-regulated (adjusted $P < 0.05$, log2FC > 1) at 72 hours AEL (compared with 120 hours AEL, which is the time point of Yki^{hyper} samples and their controls). (**D** and **E**) Average expression of the overlap of the genes down-regulated in Yki^{3SA} overexpression and *wts KO* (D) and the genes in Yki^{hyper} program (E) in the single cells of imaginal eye discs. (**F**) Expression of the Yki^{hyper} program in eye disc cell types. Heatmap (top) shows the absolute expression (log10 CPM) in single cells aggregated by cell type. Violin plot (bottom) shows the average expression in single cells grouped by cell type. (**G** to **J**) Average expression of the genes up-regulated in actively growing compared with growth-completed eye discs (G) and genes up-regulated in *Ras*^{G12V} *scrib*^{RNAi} tumors (I) and its correlation with the expression of cell cycle (H) and Yki signature (J) genes, respectively. (**K**) Overlap between the Yki^{hyper} program and genes up-regulated in actively growing compared with growth-completed eye discs and genes up-regulated in *Ras*^{G12V} *scrib*^{RNAi} tumors (adjusted $P < 0.05$, log2FC > 1). (**L**) Correlation of log2FCs in gene expression between Yki^{3SA} overexpression relative to control and growth completed discs relative to growing discs. Cell cycle and Yki signature genes are marked in blue and red, respectively, and their distribution along the axes is shown by the density plots. (**M**) Correlation between the average expression of the indicated gene sets in single cells of the eye disc. "Peripodium" refers to the marker genes (adjusted $P < 0.05$; log2FC > 0.5) of all peripodial cells. (**N**) Schematic representation of the results of the in vivo RNAi screens. (**O**) Suppressors of Yki-induced overgrowth listed in nine functional categories.

of any of them alone does not phenocopy the amount of overgrowth seen in Hippo mutants (*25*). Conversely, the 17 Yki^{1SA} overgrowth–specific suppressors included Sd (confirming that it is an overgrowth-specific gene, not a general growth gene) and several genes encoding components functioning in cytoskeleton and ECM regulation such as RhoGAP18B (a GTPase activating protein for Rho GTPases), Myosin alkali light chain 1 (Mlc1), phosphatase and actin regulator 1 (PHACTR1, CG32264), the endopeptidase Spn55B, Miles to go (Mtgo, a fibronectin type III domain containing protein), and Netrin B (NetB, a secreted ligand regulating axon outgrowth). Thus, ectopic Yki

causes eye disc overgrowth through mechanisms that include effects on the cytoskeleton and on cell-cell and cell-ECM interactions.

Discussion

Here, we evaluated the role of the Hippo pathway in organ growth control. We examined fly eye imaginal discs and the mouse liver, two commonly used model systems to study organ growth. Our data show that Hippo pathway output is not required for the proliferation of liver precursor cells and disc proper cells, the main parenchymal cell types whose proliferation embodies the growth of the liver and eye discs. We conclude that the Hippo pathway does not instruct cells when to proliferate and when to stop proliferation during the normal development of these organs.

How can this conclusion be consistent with the fact that loss of the Hippo core kinases causes severe overgrowth of liver and imaginal discs, thus demonstrating that proper activity of the kinases is required for normal organ growth? This seeming paradox is resolved when considering that the Hippo kinase mutant overgrowth phenotypes are the result of ectopic Yap/Taz/Yki activation and therefore represent gain-of-function phenotypes. Yap/Taz and Yki-Sd are normally not active in hepatocytes and disc proper cells, but their ectopic activation can induce the overexpression of a battery of genes that then cause overgrowth. This means that the role of the Hippo core kinases in hepatocytes and disc proper cells is to keep Yap/Taz/Yki in check to prevent them from triggering abnormal overgrowth. It follows that Hippo signaling is permissive but not instructive for organ growth control during the normal development of *Drosophila* eye discs and mouse liver.

So, what do Yap/Taz/Yki normally do and how does ectopic Yap/Taz/Yki induce organ overgrowth? Yap/Taz and Yki-Sd are transcriptionally active in and required for the development of cholangiocytes and peripodial cells, respectively. Others have reported that Yap/Taz/Yki-Sd are also required in endothelial cells, activated fibroblasts, lung AT1 cells, the squamous retinal pigment epithelial cells of zebrafish, and the squamous follicular cells of *Drosophila* egg chambers (*51*, *55*–*58*). These cells have in common that they are either flat cells that are under mechanical or morphological stress or that they have extra strong cell-ECM adhesion. Consistent with this theme, when we analyzed expression of the Yki-Sd reporter *kibra-lacZ*, we found strong expression in glial cells, in particular, in the two carpet cells of developing eye discs, two huge, flat cells that each covers half of the eye disc (*59*). Therefore, there is a strong correlation between cell morphology and Hippo pathway activity. Indeed, Yap/Taz are activated by mechanical stress, for example, when cells stretch out on a stiff cell culture substrate (*60*) and mediate a cell's response to mechanical strain such as the reinforcement of the cytoskeleton and cell adhesion. A major function of Yap/Taz/Yki-Sd thus appears to be a homeostatic function in morphologically and mechanically challenged cells. Consistent with this, we found that Yap and Yki hyperactivation induced factors involved in cytoskeleton and ECM regulation and activated genetic programs that were most similar to the programs of cells in which Yap/Taz/Yki-Sd are active and required for their development. We thus hypothesize that ectopic activation of these programs changes how cells interact with their neighbors and how they receive and interpret growth control signals. In this way, hyperactivation of Hippo pathway output can cause ectopic cell proliferation and organ overgrowth even though it is not directly involved in normal growth control.

Although Hippo output does not play a role in growth control in liver and eye discs, it is required for the growth of other organs. In flies, modulation of Yki is required to fine-tune the growth of the presumptive wing tissue, the wing pouch of the wing disc, where Yki is activated in peripheral cells because of the mechanical strain exerted by the growth of cells in the center (*61*). In the mouse, Yap/Taz are required for the growth of the embryonic heart, whereas hyperactivation of Yap causes heart enlargement and can trigger cardiomyocyte proliferation in adults (*62*). In these tissues, Yap/Taz/Yki appear to be activated by mechanical stress (e.g., heartbeat) and their activity may affect tissue growth by regulating the mechanical properties of the cells. Thus, the function of Yap/Taz/Yki in cardiomyocytes and wing pouch cells may be comparable to that in cholangiocytes, peripodial cells, and other mechanically challenged cells and does not reflect a role of the Hippo pathway as a master regulator of organ growth. Similarly, Yap/Taz/Yki can be activated by mechanical strain after injury and then promote tissue repair and regeneration (*2*, *3*). In the adult fly gut, for example, Yki is not required for normal homeostasis, but injury activates Yki in differentiated enterocytes, which then produce multiple cytokines and EGFR ligands that induce regenerative proliferation of intestinal stem cells (*63*). Thus, also in these cases, Yki function is permissive but not instructive for organ growth, and its effect on growth involves a specialized response that is not essential for normal development.

Our findings correct a long-standing misconception that the Hippo pathway is a master regulator of organ growth. Hippo signaling as a master regulator of growth control has provided a straightforward and convenient explanation for its function in regeneration and cancer. Our new model for how Hippo signaling affects organ growth reveals the need to re-evaluate our understanding of Yap/Taz's function as oncogenes and in regeneration. Investigating Yap/Taz as drivers of a response to mechanical strain will identify new mechanisms that promote cancer cell phenotypes and initiate regenerative responses.

Methods summary

A full description of the methods can be found in the supplementary materials. This includes full mouse and *Drosophila* genotypes and treatment protocols, descriptions of sample processing for immunofluorescence staining and RNA-seq, statistical tests, and bioinformatics analyses.

REFERENCES AND NOTES

1. Z. Wu, K. L. Guan, Hippo signaling in embryogenesis and development. *Trends Biochem. Sci.* **46**, 51–63 (2021). doi: 10.1016/j.tibs.2020.08.008; pmid: 32928629
2. Y. Zheng, D. Pan, The Hippo signaling pathway in development and disease. *Dev. Cell* **50**, 264–282 (2019). doi: 10.1016/j.devcel.2019.06.003; pmid: 31386861
3. I. M. Moya, G. Halder, Hippo-YAP/TAZ signalling in organ regeneration and regenerative medicine. *Nat. Rev. Mol. Cell Biol.* **20**, 211–226 (2019). doi: 10.1038/s41580-018-0086-y; pmid: 30546055
4. J. R. Davis, N. Tapon, Hippo signalling during development. *Development* **146**, dev167106 (2019). doi: 10.1242/dev.167106; pmid: 31527062
5. G. Halder, R. L. Johnson, Hippo signaling: Growth control and beyond. *Development* **138**, 9–22 (2011). doi: 10.1242/dev.045500; pmid: 21138973
6. J. Dong *et al.*, Elucidation of a universal size-control mechanism in *Drosophila* and mammals. *Cell* **130**, 1120–1133 (2007). doi: 10.1016/j.cell.2007.07.019; pmid: 17889654
7. D. Rogulja, C. Rauskolb, K. D. Irvine, Morphogen control of wing growth through the Fat signaling pathway. *Dev. Cell* **15**, 309–321 (2008). doi: 10.1016/j.devcel.2008.06.003; pmid: 18694569
8. T. Aegerter-Wilmsen *et al.*, Integrating force-sensing and signaling pathways in a model for the regulation of wing imaginal disc size. *Development* **139**, 3221–3231 (2012). doi: 10.1242/dev.082800; pmid: 22833127
9. K. D. Irvine, K. F. Harvey, Control of organ growth by patterning and hippo signaling in *Drosophila*. *Cold Spring Harb. Perspect. Biol.* **7**, a019224 (2015). doi: 10.1101/cshperspect.a019224; pmid: 26032720
10. I. K. Hariharan, Organ size control: Lessons from *Drosophila*. *Dev. Cell* **34**, 255–265 (2015). doi: 10.1016/j.devcel.2015.07.012; pmid: 26267393
11. M. Kango-Singh *et al.*, Shar-pei mediates cell proliferation arrest during imaginal disc growth in *Drosophila*. *Development* **129**, 5719–5730 (2002). doi: 10.1242/dev.00168; pmid: 12421711
12. N. Tapon *et al.*, salvador Promotes both cell cycle exit and apoptosis in *Drosophila* and is mutated in human cancer cell lines. *Cell* **110**, 467–478 (2002). doi: 10.1016/S0092-8674(02)00824-3; pmid: 12202036
13. K. F. Harvey, C. M. Pfleger, I. K. Hariharan, The *Drosophila* Mst ortholog, hippo, restricts growth and cell proliferation and promotes apoptosis. *Cell* **114**, 457–467 (2003). doi: 10.1016/S0092-8674(03)00557-9; pmid: 12941274
14. S. Wu, J. Huang, J. Dong, D. Pan, hippo encodes a Ste-20 family protein kinase that restricts cell proliferation and promotes apoptosis in conjunction with salvador and warts. *Cell* **114**, 445–456 (2003). doi: 10.1016/S0092-8674(03)00549-X; pmid: 12941273
15. R. S. Udan, M. Kango-Singh, R. Nolo, C. Tao, G. Halder, Hippo promotes proliferation arrest and apoptosis in the Salvador/Warts pathway. *Nat. Cell Biol.* **5**, 914–920 (2003). doi: 10.1038/ncb1050; pmid: 14502294
16. D. Zhou *et al.*, Mst1 and Mst2 maintain hepatocyte quiescence and suppress hepatocellular carcinoma development through inactivation of the Yap1 oncogene. *Cancer Cell* **16**, 425–438 (2009). doi: 10.1016/j.ccr.2009.09.026; pmid: 19878874
17. H. Song *et al.*, Mammalian Mst1 and Mst2 kinases play essential roles in organ size control and tumor suppression. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1431–1436 (2010). doi: 10.1073/pnas.0911409107; pmid: 20080598

18. F. D. Camargo *et al*., YAP1 increases organ size and expands undifferentiated progenitor cells. *Curr. Biol.* **17**, 2054–2060 (2007). doi: 10.1016/j.cub.2007.10.039; pmid: 17980593
19. L. Lu *et al*., Hippo signaling is a potent in vivo growth and tumor suppressor pathway in the mammalian liver. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1437–1442 (2010). doi: 10.1073/pnas.0911427107; pmid: 20080689
20. J. H. Lee *et al*., A crucial role of WW45 in developing epithelial tissues in the mouse. *EMBO J.* **27**, 1231–1242 (2008). doi: 10.1038/emboj.2008.63; pmid: 18369314
21. J. Huang, S. Wu, J. Barrera, K. Matthews, D. Pan, The Hippo signaling pathway coordinately regulates cell proliferation and apoptosis by inactivating Yorkie, the *Drosophila* Homolog of YAP. *Cell* **122**, 421–434 (2005). doi: 10.1016/j.cell.2005.06.007; pmid: 16096061
22. S. Wu, Y. Liu, Y. Zheng, J. Dong, D. Pan, The TEAD/TEF family protein Scalloped mediates transcriptional output of the Hippo growth-regulatory pathway. *Dev. Cell* **14**, 388–398 (2008). doi: 10.1016/j.devcel.2008.01.007; pmid: 18258486
23. E. Verboven *et al*., Regeneration defects in Yap and Taz mutant mouse livers are caused by bile duct disruption and cholestasis. *Gastroenterology* **160**, 847–862 (2021). doi: 10.1053/j.gastro.2020.10.035; pmid: 33127392
24. L. Lu, M. J. Finegold, R. L. Johnson, Hippo pathway coactivators Yap and Taz are required to coordinate mammalian liver regeneration. *Exp. Mol. Med.* **50**, e423 (2018). pmid: 29303509
25. R. M. Neto-Silva, B. S. Wells, L. A. Johnston, Mechanisms of growth and homeostasis in the *Drosophila* wing. *Annu. Rev. Cell Dev. Biol.* **25**, 197–220 (2009). doi: 10.1146/annurev.cellbio.24.110707.175242; pmid: 19575645
26. B. Z. Stanger, The biology of organ size determination. *Diabetes Obes. Metab.* **10** (Suppl 4), 16–22 (2008). doi: 10.1111/j.1463-1326.2008.00938.x; pmid: 18834429
27. M. Xin *et al*., Hippo pathway effector Yap promotes cardiac regeneration. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13839–13844 (2013). doi: 10.1073/pnas.1313192110; pmid: 23918388
28. M. Xin *et al*., Regulation of insulin-like growth factor signaling by Yap governs cardiomyocyte proliferation and embryonic heart size. *Sci. Signal.* **4**, ra70 (2011). doi: 10.1126/scisignal.2002278; pmid: 22028467
29. C. M. Weisend, J. A. Kundert, E. S. Suvorova, J. R. Prigge, E. E. Schmidt, Cre activity in fetal albCre mouse hepatocytes: Utility for developmental studies. *Genesis* **47**, 789–792 (2009). pmid: 19830819
30. C. Kellendonk, C. Opherk, K. Anlag, G. Schütz, F. Tronche, Hepatocyte-specific expression of Cre recombinase. *Genesis* **26**, 151–153 (2000). doi: 10.1002/(SICI)1526-968X(200002)26:2<151:AID-GENE17>3.0.CO;2-E; pmid: 10686615
31. M. Gordillo, T. Evans, V. Gouon-Evans, Orchestrating liver development. *Development* **142**, 2094–2108 (2015). doi: 10.1242/dev.114215; pmid: 26081571
32. N. Zhang *et al*., The Merlin/NF2 tumor suppressor functions through the YAP oncoprotein to regulate tissue homeostasis in mammals. *Dev. Cell* **19**, 27–38 (2010). doi: 10.1016/j.devcel.2010.06.015; pmid: 20643348
33. L. Zhang *et al*., The TEAD/TEF family of transcription factor Scalloped mediates Hippo signaling in organ size control. *Dev. Cell* **14**, 377–387 (2008). doi: 10.1016/j.devcel.2008.01.006; pmid: 18258485
34. Y. Goulev *et al*., SCALLOPED interacts with YORKIE, the nuclear effector of the hippo tumor-suppressor pathway in *Drosophila*. *Curr. Biol.* **18**, 435–441 (2008). doi: 10.1016/j.cub.2008.02.034; pmid: 18313299
35. L. M. Koontz *et al*., The Hippo effector Yorkie controls normal tissue growth by antagonizing scalloped-mediated default repression. *Dev. Cell* **25**, 388–401 (2013). doi: 10.1016/j.devcel.2013.04.021; pmid: 23725764
36. H. Oh, K. D. Irvine, Cooperative regulation of growth by Yorkie and Mad through bantam. *Dev. Cell* **20**, 109–122 (2011). doi: 10.1016/j.devcel.2010.12.002; pmid: 21238929
37. H. W. Peng, M. Slattery, R. S. Mann, Transcription factor choice in the Hippo signaling pathway: Homothorax and yorkie regulation of the microRNA bantam in the progenitor domain of the *Drosophila* eye imaginal disc. *Genes Dev.* **23**, 2307–2319 (2009). doi: 10.1101/gad.1820009; pmid: 19762509
38. G. Halder *et al*., The Vestigial and Scalloped proteins act together to directly regulate wing-specific gene expression in *Drosophila*. *Genes Dev.* **12**, 3900–3909 (1998). doi: 10.1101/gad.12.24.3900; pmid: 9869643
39. H. Oh, K. D. Irvine, In vivo analysis of Yorkie phosphorylation sites. *Oncogene* **28**, 1916–1927 (2009). doi: 10.1038/onc.2009.43; pmid: 19330023
40. T. Zhang, Q. Zhou, F. Pignoni, Yki/YAP, Sd/TEAD and Hth/MEIS control tissue specification in the *Drosophila* eye disc epithelium. *PLOS ONE* **6**, e22278 (2011). doi: 10.1371/journal.pone.0022278; pmid: 21811580
41. M. C. Gibson, G. Schubiger, Peripodial cells regulate proliferation and patterning of *Drosophila* imaginal discs. *Cell* **103**, 343–350 (2000). doi: 10.1016/S0092-8674(00)00125-2; pmid: 11057906
42. K. D. McClure, G. Schubiger, Developmental analysis and squamous morphogenesis of the peripodial epithelium in *Drosophila* imaginal discs. *Development* **132**, 5033–5042 (2005). doi: 10.1242/dev.02092; pmid: 16236766
43. D. M. Vallejo, S. Juarez-Carreño, J. Bolivar, J. Morante, M. Dominguez, A brain circuit that synchronizes growth and maturation revealed through Dilp8 binding to Lgr3. *Science* **350**, aac6767 (2015). doi: 10.1126/science.aac6767; pmid: 26429885
44. Y. Wang *et al*., The Cancer Genome Atlas Research Network, F. Camargo, H. Liang, Comprehensive molecular characterization of the Hippo signaling pathway in cancer. *Cell Rep.* **25**, 1304–1317.e5 (2018). doi: 10.1016/j.celrep.2018.10.001; pmid: 30380420
45. M. Guilliams *et al*., Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell* **185**, 379–396.e38 (2022). doi: 10.1016/j.cell.2021.12.018; pmid: 35021063
46. S. Manmadhan, U. Ehmer, Hippo signaling in the live – A long and ever-expanding story. *Front. Cell Dev. Biol.* **7**, 33 (2019). doi: 10.3389/fcell.2019.00033; pmid: 30931304
47. L. Yang *et al*., A single-cell transcriptomic analysis reveals precise pathways and regulatory mechanisms underlying hepatoblast differentiation. *Hepatology* **66**, 1387–1401 (2017). doi: 10.1002/hep.29353; pmid: 28681484
48. T. Katsuyama, R. Paro, Innate immune cells are dispensable for regenerative growth of imaginal discs. *Mech. Dev.* **130**, 112–121 (2013). doi: 10.1016/j.mod.2012.11.005; pmid: 23238120
49. K. Strassburger, M. Lutz, S. Müller, A. A. Teleman, Ecdysone regulates *Drosophila* wing disc size via a TORC1 dependent mechanism. *Nat. Commun.* **12**, 6684 (2021). doi: 10.1038/s41467-021-26780-0; pmid: 34795214
50. C. Bravo González-Blas *et al*., Identification of genomic enhancers through spatial integration of single-cell transcriptomics and epigenomics. *Mol. Syst. Biol.* **16**, e9438 (2020). doi: 10.15252/msb.20209438; pmid: 32431014
51. G. C. Fletcher *et al*., Mechanical strain regulates the Hippo pathway in *Drosophila. Development* **145**, dev.159467 (2018). doi: 10.1242/dev.159467; pmid: 29440303
52. A. Genevet, M. C. Wehr, R. Brain, B. J. Thompson, N. Tapon, Kibra is a regulator of the Salvador/Warts/Hippo signaling network. *Dev. Cell* **18**, 300–308 (2010). doi: 10.1016/j.devcel.2009.12.011; pmid: 20159599
53. M. Torres-Oliva, J. Schneider, G. Wiegleb, F. Kaufholz, N. Posnien, Dynamic genome wide expression profiling of *Drosophila* head development reveals a novel role of Hunchback in retinal glia cell development and blood-brain barrier integrity. *PLOS Genet.* **14**, e1007180 (2018). doi: 10.1371/journal.pgen.1007180; pmid: 29360820
54. M. Atkins *et al*., An ectopic network of transcription factors regulated by hippo signaling drives growth and invasion of a malignant tumor model. *Curr. Biol.* **26**, 2101–2113 (2016). doi: 10.1016/j.cub.2016.06.035; pmid: 27476594
55. X. Wang *et al*., YAP/TAZ orchestrate VEGF signaling during developmental angiogenesis. *Dev. Cell* **42**, 462–478.e7 (2017). doi: 10.1016/j.devcel.2017.08.002; pmid: 28867486
56. I. Mannaerts *et al*., The Hippo pathway effector YAP controls mouse hepatic stellate cell activation. *J. Hepatol.* **63**, 679–688 (2015). doi: 10.1016/j.jhep.2015.04.011; pmid: 25908270
57. J. B. Miesfeld *et al*., Yap and Taz regulate retinal pigment epithelial cell fate. *Development* **142**, 3021–3032 (2015). pmid: 26209646
58. L. B. Nantie *et al*., *Lats1/2* inactivation reveals Hippo function in alveolar type I cell differentiation during lung transition to air breathing. *Development* **145**, 10.1242/dev.163105 (2018). pmid: 30305289
59. M. Silies *et al*., Glial cell migration in the eye disc. *J. Neurosci.* **27**, 13130–13139 (2007). doi: 10.1523/JNEUROSCI.3583-07.2007; pmid: 18045907
60. S. Dupont *et al*., Role of YAP/TAZ in mechanotransduction. *Nature* **474**, 179–183 (2011). doi: 10.1038/nature10137; pmid: 21654799
61. C. Rauskolb, S. Sun, G. Sun, Y. Pan, K. D. Irvine, Cytoskeletal tension inhibits Hippo signaling through an Ajuba-Warts complex. *Cell* **158**, 143–156 (2014). doi: 10.1016/j.cell.2014.05.035; pmid: 24995985
62. S. Liu, J. F. Martin, The regulation and function of the Hippo pathway in heart regeneration. *Wiley Interdiscip. Rev. Dev. Biol.* **8**, e335 (2019). doi: 10.1002/wdev.335; pmid: 30169913
63. P. Zhang, B. A. Edgar, Insect gut regeneration. *Cold Spring Harb. Perspect. Biol.* **14**, a040915 (2022). doi: 10.1101/cshperspect.a040915; pmid: 34312250

ACKNOWLEDGMENTS

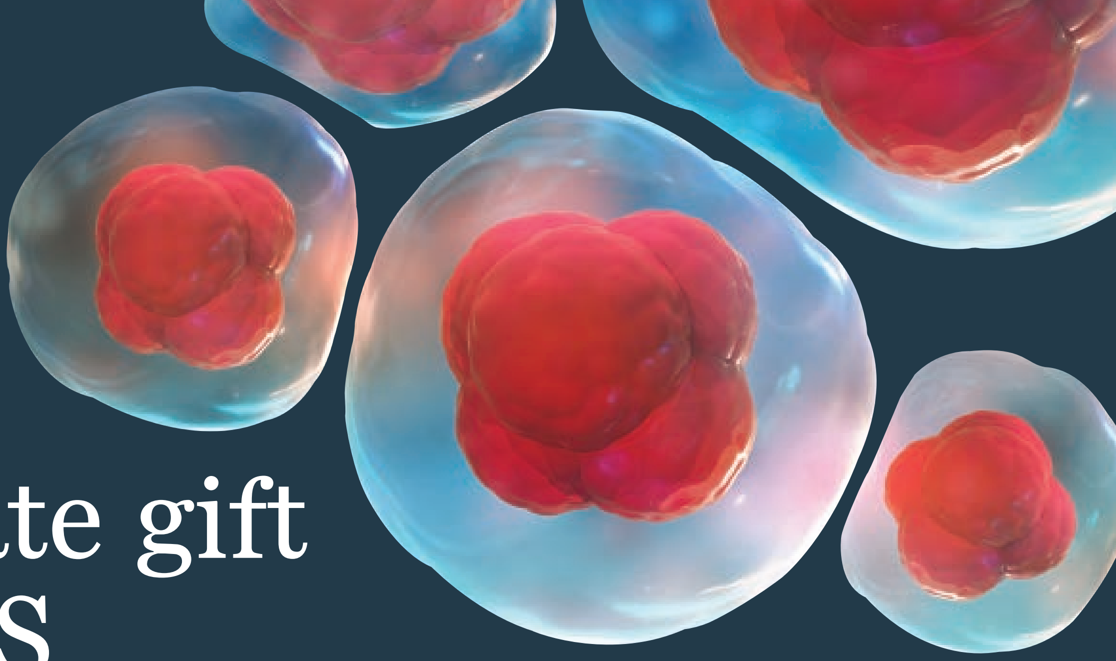
We thank D. Pan for providing the *sd*, *hpo*, and *yki* mutant fly stocks and *ApoE>hYAP* mice; K. D. Irvine for the yki^{3SA} stock; H. Bellen for the gp-anti-Ato antibody; the Transgenic RNAi Project (TRiP) at Harvard Medical School, the Vienna *Drosophila* Resource Center, and the Bloomington *Drosophila* Stock Center for providing transgenic RNAi and other fly stocks; and N. Alves Souza Carvalhais, E. Kesikiadou, and J. Mach for help with experiments. **Funding:** This work was supported by The Research Foundation Flanders (FWO Odysseus type I grant and project grants G.0954.16N and G.030616N to G.H. and an FWO doctoral fellowship to W.K.) and by the Cancer Prevention and Research Institute of Texas (grant RP200240 to R.J.). **Author contributions:** G.H. conceived the project. G.H., W.K., L.R., I.M.M., and M.A. wrote the manuscript. G.H., W.K., L.R., and M.A. designed the experiments and analyzed the data. W.K. and L.R. organized and performed most of the experiments. M.A. contributed to *Drosophila* experiments. W.K., C.B.G.-B., J.J., M.d.W., and S.A. performed the bioinformatics analyses. H.H., I.M.M., E.V., S.S., S.V., L.S.-G., L.v.H., R.J., and L.A.v.G. contributed to the mouse experiments. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All sequencing data have been deposited in the Gene Expression Omnibus (GEO) under the accession number GSE211461 with the following subseries: GSE211457 (*Drosophila* – nonpupating larvae), GSE211458 (*Drosophila* – Yki hyperactivation), GSE211459 (liver – Yap hyperactivation), and GSE211460 (liver – CCl_4 in Yap/Taz KO). Processed data and analyses results are provided in the supplementary tables.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg3679
Materials and Methods
Figs. S1 to S17
Tables S1 to S10
References (*64*–*79*)
MDAR Reproducibility Checklist

Submitted 31 December 2020; resubmitted 9 August 2022
Accepted 12 October 2022
10.1126/science.abg3679



An estate gift to AAAS

Going all the way back to 1848, our founding year, the American Association for the Advancement of Science (AAAS) has been deeply committed to advancing science, engineering and innovation around the world for the benefit of all people.

Today, we are dedicated to advocating for science and scientific evidence to be fully and positively integrated into public policy and for the community to speak with one voice to advance science and engineering in the United States and around the world.

By making AAAS a beneficiary of your will, trust, retirement plan or life insurance policy, you will become a member of our 1848 Society and will help fuel our work on behalf of science and society – including publishing the world's most promising, innovative research in the *Science* family of journals and engaging in the issues that matter locally, nationally and around the world.

"As a teacher and instructor, I bear responsibility for the younger generations. If you have extra resources, concentrate them on organizations, like AAAS, that are doing work for all."

—Prof. Elisabeth Ervin-Blankenheim, 1848 Society member

If you intend to include AAAS in your estate plans, provide this information to your lawyer or financial adviser:

Legal Name: American Association for the Advancement of Science
Federal Tax ID Number: 53-0196568
Address: 1200 New York Avenue, NW, Washington, DC 20005

If you would like more information on making an estate gift to AAAS, cut out and return the form below or send an email to philanthropy@aaas.org. Additional details are also available online at **www.aaas.org/1848Society**.

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

cut here

Yes, I would like more information about joining the AAAS 1848 Society.

PLEASE CONTACT ME AT:

Name: ______________________

Address: ______________________

City: ____________ State: ______ Zip code: __________ Country: ________

Email: ______________ Phone: ______________

RETURN THIS FORM TO:

AAAS Office of Philanthropy and Strategic Partnerships • 1200 New York Avenue, NW • Washington, DC 20005 USA

RESEARCH ARTICLE SUMMARY

HUMAN FERTILITY

The mechanism of acentrosomal spindle assembly in human oocytes

Tianyu Wu†, Jie Dong†, Jing Fu†, Yanping Kuang†, Biaobang Chen, Hao Gu, Yuxi Luo, Ruihuan Gu, Meiling Zhang, Wen Li, Xi Dong, Xiaoxi Sun*, Qing Sang*, Lei Wang*

INTRODUCTION: Spindle assembly is essential for ensuring accurate chromosome transmission in both meiosis and mitosis. In somatic cells, mitotic spindle assembly is mediated by duplicated centrosomes, but canonical centrosomes are absent in the oocytes of many species. In rodents, acentriolar microtubule organizing centers (aMTOCs) are responsible for meiotic spindle assembly, but it has long been supposed that human oocytes lack prominent aMTOCs on the meiotic spindle, and the exact mechanism of acentrosomal spindle assembly in human oocytes has remained unclear.

RATIONALE: Microtubule nucleation and ensuring spindle assembly are core events regulating oocyte nuclear maturation. To identify the potential proteins driving spindle microtubule nucleation in human oocytes, we systematically localized 86 human centrosome and microtubule-related proteins by immunofluorescence or three-dimensional high-resolution live cell imaging in more than 2000 human oocytes. We then tracked the dynamic migration of identified microtubule nucleators at different time points before and after nuclear envelope breakdown (NEBD). We further downregulated corresponding proteins to confirm their role in microtubule nucleation and spindle assembly. Given that spindle microtubule nucleation defects result in impaired spindle assembly and abnormal oocyte maturation, we screened for mutations in genes encoding components of microtubule nucleators in a cohort of 1394 infertile female patients characterized by oocyte maturation arrest.

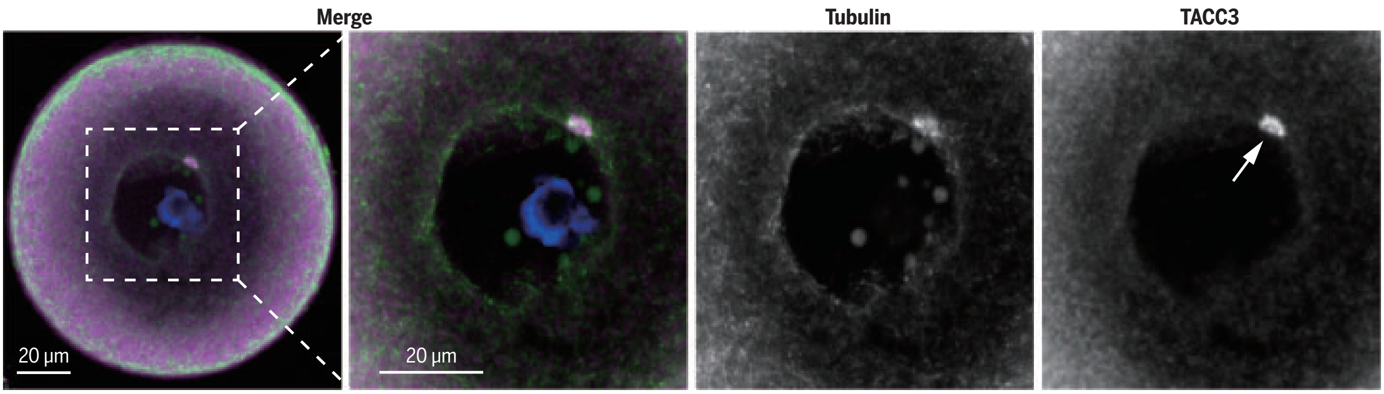
RESULTS: First, we found that in human oocytes the nucleation of spindle microtubules is initiated from kinetochores from 2 to 4 hours after NEBD. We showed the process of spindle microtubules nucleating from kinetochores in human oocytes. We then found that there are 43 proteins localized in the meiotic spindle, among which four proteins—centriolar coiled-coil protein 110 (CCP110), cytoskeleton-associated protein 5 (CKAP5), disrupted in schizophrenia 1 (DISC1), and transforming acidic coiled-coil–containing protein 3 (TACC3)—exhibited both kinetochore and spindle microtubule localization. The localization of the four proteins was notably different from their localization in human mitotic cells and in mouse oocytes. Together, the four proteins formed an unusual structure that was surrounded by microtubules in human germinal vesicle (GV) oocytes just before NEBD. We refer to this potential nucleating structure as the human oocyte microtubule organizing center (huoMTOC). We found that a single huoMTOC is formed at the cortex of human GV oocytes and migrates to the nuclear envelope before NEBD. After NEBD, the huoMTOC becomes fragmented and is recruited to kinetochores to initiate spindle microtubule nucleation. Down-regulation of huoMTOC components caused considerably impaired spindle microtubule nucleation and spindle assembly in human oocytes. This structure was not detected in the oocytes of other mammalian species such as mice and pigs. We finally identified two oocyte maturation arrest patients with compound heterozygous mutations in the key huoMTOC component *TACC3*. All mutations disrupted the normal function of TACC3, resulting in the absence of the huoMTOC structure and completely impaired spindle assembly in the patients' oocytes.

CONCLUSION: Our study shows that human oocytes possess an aMTOC-like structure, the huoMTOC, that serves as a major site of microtubule nucleation and is required for spindle assembly. The huoMTOC shows drastically different characteristics in terms of number, localization, and composition compared with aMTOCs in mouse oocytes. These findings suggest that a distinct mechanism for the initiation of microtubule nucleation and spindle assembly has evolved in human oocytes. We found that mutations in *TACC3* cause defects in spindle assembly by disrupting the structure of the huoMTOC, which leads to clinical oocyte maturation arrest. This suggests that the huoMTOC might be an important biomarker for evaluating the quality of human oocytes.

Our discovery of huoMTOC provides insights into the physiological mechanism of microtubule nucleation and spindle assembly in human oocytes. These findings also improve our understanding of the pathological mechanisms of oocyte maturation arrest. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: wangleiwanglei@fudan.edu.cn (L.W.); sangqing@fudan.edu.cn (Q.S.); xiaoxi_sun@aliyun.com (X.S.)
†These authors contributed equally to this work.
Cite this article as T. Wu *et al.*, *Science* **378**, eabq7361 (2022). DOI: 10.1126/science.abq7361

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abq7361



The huoMTOC structure in a human oocyte. The human GV oocyte shown here was matured for ~5 hours and fixed for immunofluorescence before NEBD. The huoMTOC (TACC3, magenta) was surrounded by numerous microtubules (green) on the nuclear envelope. The dashed square shows the magnification region. The arrow highlights the huoMTOC.

RESEARCH ARTICLE

HUMAN FERTILITY

The mechanism of acentrosomal spindle assembly in human oocytes

Tianyu Wu[1]†, Jie Dong[1]†, Jing Fu[2]†, Yanping Kuang[3]†, Biaobang Chen[4], Hao Gu[1], Yuxi Luo[1], Ruihuan Gu[2], Meiling Zhang[5], Wen Li[5], Xi Dong[6], Xiaoxi Sun[2]*, Qing Sang[1]*, Lei Wang[1]*

Meiotic spindle assembly ensures proper chromosome segregation in oocytes. However, the mechanisms behind spindle assembly in human oocytes remain largely unknown. We used three-dimensional high-resolution imaging of more than 2000 human oocytes to identify a structure that we named the human oocyte microtubule organizing center (huoMTOC). The proteins TACC3, CCP110, CKAP5, and DISC1 were found to be essential components of the huoMTOC. The huoMTOC arises beneath the oocyte cortex and migrates adjacent to the nuclear envelope before nuclear envelope breakdown (NEBD). After NEBD, the huoMTOC fragments and relocates on the kinetochores to initiate microtubule nucleation and spindle assembly. Disrupting the huoMTOC led to spindle assembly defects and oocyte maturation arrest. These results reveal a physiological mechanism of huoMTOC-regulated spindle assembly in human oocytes.

Spindle assembly is essential for ensuring accurate chromosome transmission in both mitosis and meiosis (*1*, *2*). In somatic cells, mitotic spindle assembly is mediated via duplicated centrosomes, which consist of two centrioles surrounded by the pericentriolar material (PCM) (*3*, *4*). Centrosomes are the major microtubule organizing centers (MTOCs) in mitotic cells, and they are responsible for microtubule nucleation and spindle pole organization of the centrosomal spindle (*5*, *6*).

Unlike somatic cells, canonical centrosomes are absent in the oocytes of many species (*7*–*11*). Instead, acentriolar MTOCs (aMTOCs) are observed in frog (*12*) and mouse (*13*–*15*) oocytes, whereas no MTOC structures are found in the *Drosophila* (*16*) or *Caenorhabditis elegans* (*17*) oocytes. The difference suggests that the mechanisms of female meiotic spindle assembly are not conserved between species. The aMTOC-directed meiotic spindle assembly was only elucidated in mouse oocytes (*15*, *18*, *19*).

[1]Institute of Pediatrics, Children's Hospital of Fudan University, State Key Laboratory of Genetic Engineering, Institutes of Biomedical Sciences, Shanghai Key Laboratory of Medical Epigenetics, Fudan University, Shanghai 200032, China. [2]Shanghai Ji Ai Genetics and IVF Institute, Obstetrics and Gynecology Hospital, Fudan University, Shanghai 200011, China. [3]Department of Assisted Reproduction, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200011, China. [4]NHC Key Lab of Reproduction Regulation, Shanghai Institute for Biomedical and Pharmaceutical Technologies, Fudan University, Shanghai 200032, China. [5]Center for Reproductive Medicine and Fertility Preservation Program, International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200030, China. [6]Reproductive Medicine Center, Zhongshan Hospital, Fudan University, Shanghai 200032, China.
***Corresponding author. Email: wangleiwanglei@fudan.edu.cn (L.W.); sangqing@fudan.edu.cn (Q.S.); xiaoxi_sun@aliyun.com (X.S.)**
†These authors contributed equally to this work.

Mouse aMTOCs lack centrioles but contain partial centrosomal proteins such as pericentrin (*14*), γ-tubulin (*20*), CEP192 (*18*), and NEDD1 (*21*). Upon meiotic resumption after prophase I arrest, multiple aMTOCs initiate microtubule nucleation around the nuclear envelope in mouse germinal vesicle (GV) oocytes. After nuclear envelope breakdown (NEBD), aMTOCs are clustered and then concentrated at the spindle poles for bipolar spindle organization (*15*, *18*, *19*).

A mechanistic understanding of meiotic spindle assembly in human oocytes, on the other hand, remains elusive (*2*). It is only known that the spindle microtubule nucleation in human oocytes is mediated by chromosomes and promoted by guanosine triphosphate (GTP)–bound Ran (RanGTP) (*22*). Subsequently, a multipolar spindle is assembled as an intermediate, and then the spindle poles are merged to form a bipolar spindle (*22*). However, human oocytes lack detectable aMTOCs at the meiotic spindle poles (*22*, *23*), and thus the exact mechanism of acentrosomal spindle assembly in human oocytes remains unclear.

Results

Spindle microtubules are nucleated from kinetochores in human oocytes

A previous study implied that spindle microtubules emanate from the kinetochores in human oocytes (*22*), and this phenomenon was also observed in our immunofluorescent results in early prometaphase oocytes (fig. S1). This suggests that kinetochores may serve as the microtubule nucleation sites in human oocytes. However, the dynamic process of how spindle microtubules were originally nucleated from kinetochores was not delineated. Thus, we first observed the process of microtubule nucleation in live human oocytes by using three-dimensional (3D) high-resolution time-lapse imaging (Fig. 1A). Human GV oocytes were co-injected with mRNA encoding fluorescently fused histone H2B (mCherry) and centromere protein B (CENPB) (mClover3) to visualize chromosomes and kinetochores, respectively. Combined with fluorescent proteins, SiR-tubulin (a dye with far-red fluorescence) was used to label microtubules before oocyte maturation (*24*). According to our three-channel time-lapse images, a microtubule cluster (dashed square) derived from a GV oocyte was observed proximal to chromosomes upon NEBD and disassembled in the first few hours of meiosis I (Fig. 1A). Apart from the microtubules (light-gray curve, bottom panel) derived from the GV oocyte, a few nascent microtubules (dark-gray curve, bottom panel) were detected by microscopy after NEBD (Fig. 1A).

To confirm that the nascent microtubules were polymerized from kinetochores, we evaluated the tubulin intensity around kinetochores over time after NEBD. The intensity of microtubules (tubulin) and chromosomes (H2B) was measured around the representative kinetochores (Fig. 1A, arrows). The nascent microtubules (dark-gray curve) were hardly observed at the beginning of meiosis in human oocytes (0 to 2 hours after NEBD). With the disassembly of the derived microtubule cluster (light-gray curve), the fluorescent tubulins began to concentrate at the kinetochores starting at ~2 hours after NEBD (Fig. 1, A and B). According to three-channel time-lapse imaging, the fluorescence of nascent microtubules (dark-gray curve) overlapped primarily with kinetochores (red curve) from 2 to 2.5 hours after NEBD and consistently throughout the beginning of meiosis I (2 to 4 hours after NEBD) (Fig. 1A). However, the fluorescence of nascent microtubules (dark-gray curve) did not consistently overlap with chromosomes (blue curve), suggesting that the nascent microtubules primarily polymerized from the kinetochores rather than other regions of the chromosomes (Fig. 1A). Along with meiotic maturation, microtubules were nucleated slowly but continuously (Fig. 1, A and B), and the stable microtubules with high tubulin intensity could be observed on kinetochores starting at 3 hours after NEBD (Fig. 1, A and B). These results suggest that in human oocytes, the nucleation of spindle microtubules is initiated from kinetochores starting at 2 to 4 hours after NEBD.

Next, to further confirm the nucleation of spindle microtubules on kinetochores, we treated live human metaphase I (MI) oocytes with a reversible microtubule inhibitor (nocodazole) and tracked the dynamic recovery of microtubule nucleation immediately after nocodazole was washed out (Fig. 1C). Microtubules and kinetochores were marked by

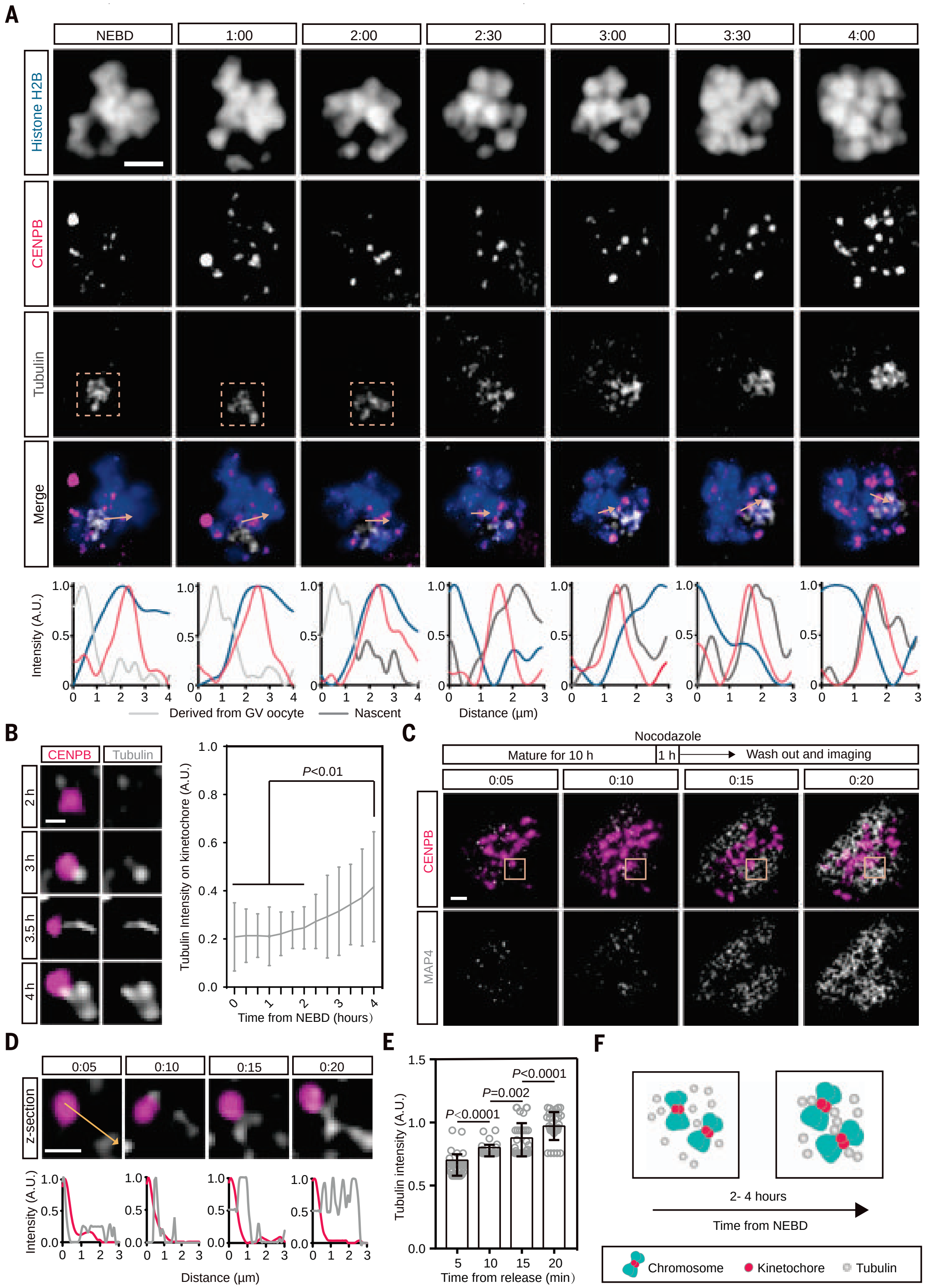


Fig. 1. The dynamic process of spindle microtubules nucleating from kinetochores in human oocytes. (**A**) Representative time-lapse images showing microtubule nucleation in human oocytes after NEBD. Blue, chromosomes (H2B-mCherry); gray, microtubules (SiR-tubulin); magenta, kinetochores (mClover-CENPB). The microtubule cluster was marked at 0 to 2 hours after NEBD (dashed squares). Representative kinetochores of each time point are shown in some z-sections. The intensity of chromosomes and microtubules was measured and plotted around representative kinetochores. The graphs in the bottom panel are the fluorescence profiles of H2B, tubulin, and CENPB across the representative kinetochores along the direction of the arrows in the merged images at each time point. The microtubules derived from GV oocytes (light-gray curves) and nascent microtubules (dark-gray curves) were distinguished at 2 hours after NEBD. The plotted intensity curves indicate the relative location of each fluorescent protein. Blue, chromosomes; gray, microtubules; magenta, kinetochores. A.U., arbitrary units. Scale bar, 5 μm. (**B**) Representative images showing kinetochore nucleated microtubules at different time points. The intensity of the microtubules on kinetochores measured in (A) was compared between 2 and 4 hours after NEBD ($n > 24$ kinetochores, significance of the differences in intensity was calculated by ANOVA and indicated on the graph). Scale bar, 2 μm. (**C**) Representative time-lapse images (z-stack) showing microtubule nucleation after nocodazole washout in live human oocytes. Gray, microtubules (GFP-MAP4); magenta, kinetochores (mScarlet-CENPB). Flow diagram shows the sequence of experiments. Arrow indicates time-lapse imaging initiation. Time is given as hours:minutes after nocodazole washout. Scale bar, 5 μm. (**D**) Single-slice images of boxed areas in (C). Time is given as hours:minutes after nocodazole washout. Scale bar, 2 μm. Intensity of kinetochores and microtubules is indicated in the graphs. The bottom graphs are the fluorescence profiles of MAP4 (gray) and CENPB (magenta) proximal to the representative kinetochores along the direction of the yellow arrow. (**E**) The intensity of microtubules on kinetochores in (D) was measured and compared after nocodazole washout (significance was calculated by ANOVA). (**F**) Mechanistic model for microtubule nucleation in human oocytes. Time is after NEBD.

fluorescent microtubule-associated protein 4 (MAP4) and CENPB proteins, respectively. Initially, nocodazole completely disrupted the microtubules in human MI oocytes, and repolymerized microtubules were not detected on kinetochores until ~10 min after nocodazole was washed out. At that time, several nascent microtubules were observed near the kinetochores, and some of them were emanating outward (Fig. 1C). The trajectory of microtubules was also shown in a single slice (Fig. 1D), and the representative kinetochore was nucleating

microtubules slowly and continuously, suggesting that nascent microtubules were nucleated from kinetochores (Fig. 1, D and E). Together, these results show the dynamic process of acentrosomal spindle microtubules nucleating from kinetochores in human oocytes (Fig. 1F).

Discovery of a specific microtubule nucleator in human oocytes

To identify the specific factors driving spindle microtubule nucleation from kinetochores in human oocytes, we localized 86 human centrosome and microtubule-related proteins by performing immunofluorescence in >1000 fixed human oocytes (Fig. 2A and figs. S2 and S3). These proteins were classified according to their function or localization in somatic cells (*21*, *25*, *26*), but their localization in human oocytes was largely unknown. These proteins included 34 centrosomal proteins, 25 microtubule-associated proteins, 12 dynein-related proteins, five regulatory kinases or substrates, five spindle assembly factors, and five nuclear pore–related proteins (fig. S2). A total of 36 proteins were specifically localized on spindle microtubules, two proteins were concentrated on spindle poles, and one protein showed spindle periphery localization (Fig. 2A and fig. S3). Unexpectedly, four proteins—centriolar coiled-coil protein 110 (CCP110), cytoskeleton-associated protein 5 (CKAP5), disrupted in schizophrenia 1 (DISC1), and transforming acidic coiled-coil–containing protein 3 (TACC3)—exhibited both kinetochore and spindle microtubule localization (Fig. 2B), and such localization was consistent until the MII stage (fig. S4), which was notably different from their localization in human mitotic cells and in mouse oocytes (*6*, *21*, *27*–*30*).

CCP110 and DISC1 are centrosomal proteins that are involved in spindle assembly at the centrosomes of human mitotic cells and aMTOCs of mouse oocytes (*6*, *21*, *27*). CKAP5 and TACC3 are microtubule-associated proteins concentrated at centrosomes and microtubules in human mitotic cells during mitosis (*28*–*30*). Importantly, both CKAP5 and TACC3 have been reported to nucleate microtubules in vitro and in mitotic cells, respectively (*31*–*34*). These observations suggested that CCP110, CKAP5, DISC1, and TACC3 are potential

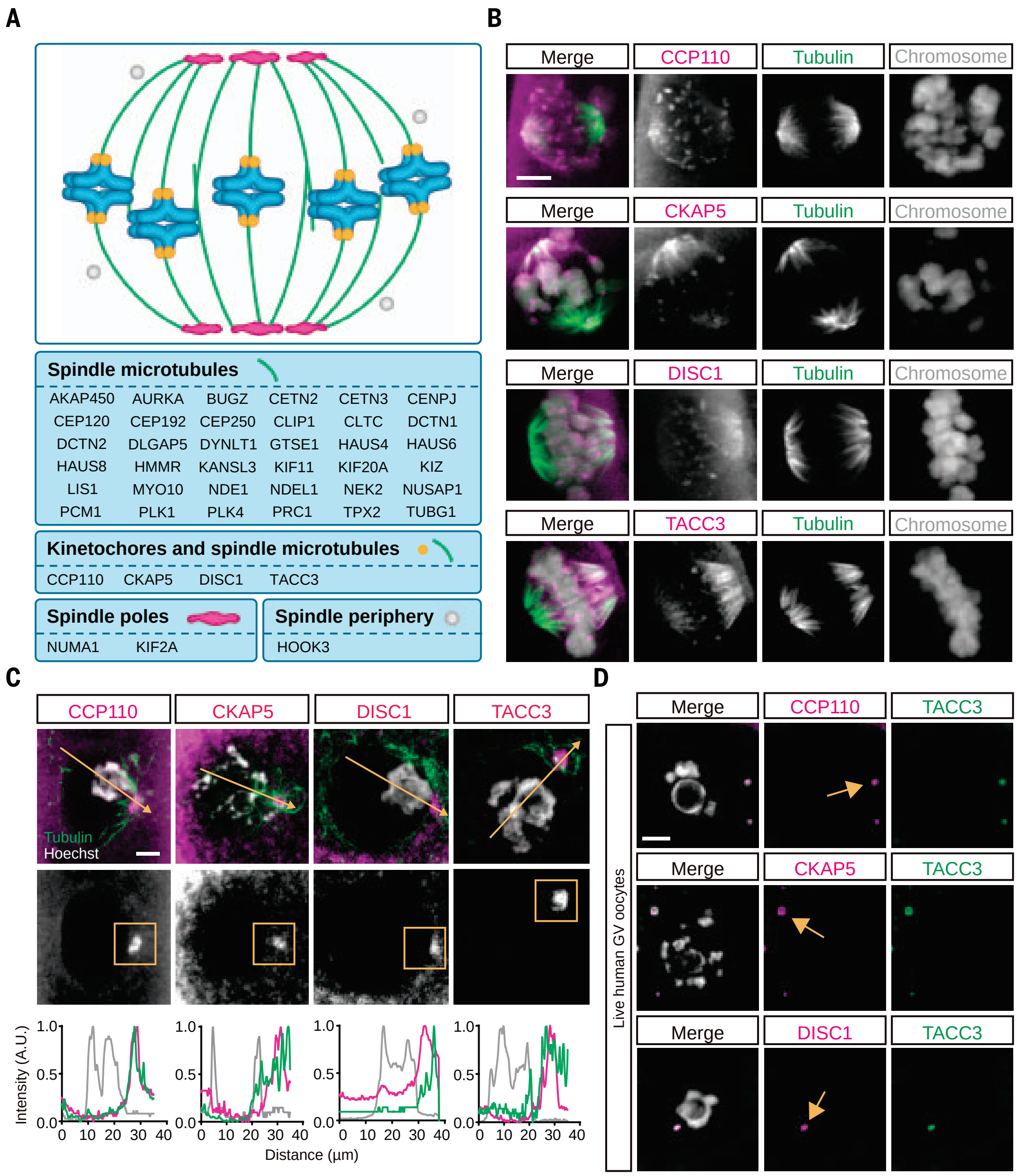


Fig. 2. Identification of a microtubule nucleator in human oocytes. (**A**) Schematic of the metaphase I spindle in human oocytes (chromosomes in blue, microtubules in green, spindle poles in magenta, kinetochores in yellow, and spindle periphery in gray). (**B**) Immunofluorescence images of metaphase I spindle in human MI oocytes showing protein localization relative to both microtubules and chromosomes. Gray, chromosomes; green, microtubules. Scale bar, 5 μm. (**C**) Immunofluorescence images of the nucleus in human GV oocytes before NEBD. The yellow squares indicate the newly identified structure surrounded by microtubules. Scale bar, 10 μm. The yellow arrows indicate intensity measurements in the images at the top. The localization of huoMTOC and microtubule distribution of human GV oocytes are shown below. Gray, chromatin (Hoechst); green, microtubules (tubulin); magenta, huoMTOC. (**D**) Representative images of live human GV oocytes. mScarlet-TACC3 was used as the standard for colocalization. Green, TACC3; magenta, CCP110, CKAP5, or DISC1; gray, chromosomes (Hoechst). The yellow arrows indicate the colocalization. Scale bar, 10 μm.

candidates for regulating microtubule nucleation and polymerization in human oocytes.

Subsequently, these proteins were monitored by immunofluorescence in human GV oocytes just before NEBD (fig. S5 and movie S1). Notably, each of the four proteins showed an unusual structure surrounded by microtubules. This structure was proximal to the nuclear envelope in human GV oocytes ~0 to 2 hours before NEBD (Fig. 2C), which was consistent with our observations of a microtubule cluster proximate to chromatin in human oocytes after NEBD (movie S2). To test whether these four proteins were colocalized, we overexpressed TACC3 (labeled with mScarlet) and the other three proteins (labeled with mClover3). As indicated in Fig. 2D, CCP110, CKAP5, and DISC1 were all colocalized with TACC3, implying that the four proteins belong to the same structure. To examine the interactions among these proteins, coimmunoprecipitation was performed in human embryonic kidney 293T (HEK293T) cells transfected with plasmids containing the corresponding genes. As a result, CCP110, CKAP5, and DISC1 all interacted directly with TACC3, whereas the negative control GDF9 had no interaction with TACC3 (fig. S6). This suggests that the four proteins are all components of the same structure. In addition, a dense microtubule cluster was observed around this structure in human GV oocytes, which caused asymmetrical microtubule distribution around the nuclear envelope prior to NEBD (Fig. 2C). Given these features, we refer to this potential nucleating structure as the human oocyte microtubule organizing center (huoMTOC). This structure was not detected in the oocytes of other mammalian species such as mice and pigs (fig. S7), suggesting the specificity of this structure in human oocytes.

The huoMTOC is essential for microtubule nucleation

According to the transcriptional landscape data of human oocytes in public databases (*23*), *TACC3* shows overwhelmingly higher expression (more than 29-fold) than *CCP110*, *CKAP5*, and *DISC1*, implying its potential key role in the huoMTOC. We therefore tried to disrupt the huoMTOC by knocking down *TACC3* by injecting the corresponding short-interfering RNAs (siRNAs) into human GV oocytes (fig. S8). The integrity of the huoMTOC was evaluated by immunofluorescence for CCP110, CKAP5, DISC1, and TACC3. As expected, no nucleating structures were detected, demonstrating the complete disruption of the huoMTOC upon TACC3 depletion (Fig. 3A). To determine whether huoMTOC is the main microtubule nucleator, we assessed the microtubule distribution in oocytes under the condition of huoMTOC deficiency. The asymmetrical distribution of microtubules was drastically diminished, implying the essential role of the huoMTOC for microtubule nucleation in human oocytes (Fig. 3A). In addition, different degrees of disruption in the asymmetrical distribution of microtubules around the nuclear envelope were also observed after specific depletion of endogenous CCP110, CKAP5, or DISC1 (Fig. 3, B and C, and fig. S8). Of note, depletion of TACC3 resulted in the most severe disruption of microtubule asymmetrical distribution in most of the analyzed human GV oocytes (Fig. 3, B and C). These results indicate that the huoMTOC is essential for the asymmetrical nucleation of microtubules around the nuclear envelope and that all four proteins are indispensable for maintaining normal function of the huoMTOC, in which TACC3 presumably plays a leading role.

The huoMTOC is fragmented and recruited to kinetochores for the initiation of spindle assembly in human oocytes

Unlike the mechanism in mouse MI oocytes in which aMTOCs were aggregated on the spindle poles during spindle assembly (*15*), the components of the huoMTOC in human MI oocytes were localized on kinetochores. To reveal the dynamic process of huoMTOC-regulated spindle assembly, human GV oocytes were fixed for immunofluorescence at NEBD or at 2, 4, or 6 hours before NEBD. Initially, the huoMTOC appeared beneath the oocyte cortex. Slowly, the huoMTOC migrated from the cortex to the nuclear envelope of human GV oocytes (Fig. 4, A and B). The huoMTOC then expanded, and its nucleated microtubules grew rapidly during the resumption of meiosis (Fig. 4, A and C).

We next visualized the dynamic changes of the huoMTOC after NEBD by 3D time-lapse imaging. At the beginning of NEBD, the huoMTOC localized proximal to the chromosomes and became fragmented within the first hour (fig. S9A and movie S3). The huoMTOC microtubules were disrupted and were barely observable in the first few hours after NEBD (Fig. 4D and movie S4). The fragmented huoMTOC was then relocated to chromosomes, primarily to kinetochores (Fig. 4D, fig. S9B, and movie S5). The spindle microtubules were initially observed when the huoMTOC was rebuilt (Fig. 4, D and E; fig. S9C; and movie S4).

Next, to determine the role of the huoMTOC in spindle assembly, each huoMTOC component was down-regulated in human GV oocytes that were cultured to the MI stage. Down-regulation of these components significantly impaired spindle microtubule nucleation and spindle assembly in human MI oocytes compared with the control group ($P < 0.001$, Fisher's exact test), and stable microtubules were greatly decreased (Fig. 4, F and G). Compared to other components, TACC3 depletion had the most obvious effects on the microtubule nucleation and spindle assembly (Fig. 4, F and G), further suggesting that TACC3 plays a leading role in the huoMTOC. To directly test whether huoMTOC is essential for spindle assembly, the huoMTOC marked by fluorescent TACC3 was disrupted by laser ablation (fig. S10, A and B). Similar to TACC3 depletion, the spindle microtubule polymerization and spindle assembly were significantly impaired by the laser ablation of huoMTOC ($P = 0.015$, Fisher's exact test) (fig. S10, C and D). It has been demonstrated that RanGTP is also required for spindle assembly in human oocytes (*22*). Disruption of both TACC3 and RanGTP aggravated the spindle microtubule polymerization defects (fig. S11), indicating combined effects of TACC3 and RanGTP on microtubule nucleation and spindle assembly. These results suggest that the huoMTOC is required for spindle microtubule polymerization and spindle assembly of human oocytes.

huoMTOC deficiency interrupts normal spindle assembly and causes clinical oocyte maturation arrest

Oocyte maturation requires microtubule nucleation and spindle assembly (*22*). In the clinic, a number of infertile patients with recurrent failed in vitro fertilization (IVF) or intracytoplasmic sperm injection (ICSI) attempts have been diagnosed with oocyte maturation arrest. Considering the key role of the huoMTOC in human spindle assembly, we hypothesized that a disrupted huoMTOC resulting from mutations in *CCP110*, *CKAP5*, *DISC1*, or *TACC3* may cause impaired spindle assembly and abnormal oocyte maturation in patients. We thus screened for likely pathogenic mutations in a cohort of 1394 infertile female patients characterized by oocyte maturation arrest by analyzing their whole exome sequencing datasets (data deposited in the Genome Variation Map of the National Genomics Data Center under accession number GVM000402). Each of these patients had undergone several IVF attempts, all of which failed because of oocyte maturation arrest. We identified two patients with compound heterozygous mutations in the key huoMTOC component *TACC3* (Fig. 5A and table S1). According to the Genome Aggregation Database (gnomAD) and our in-house controls (data deposited in the Genome Variation Map of the National Genomics Data Center under accession number GVM000394), all four *TACC3* variants from patients are rare variants (table S2). Importantly, these two patients showed very similar phenotypes, in which most of their retrieved oocytes were immature after in vitro maturation (table S1), and polarization microscopy images showed no visible spindles in the live oocytes (Fig. 5B). Immunofluorescence in fixed NEBD oocytes also demonstrated that the

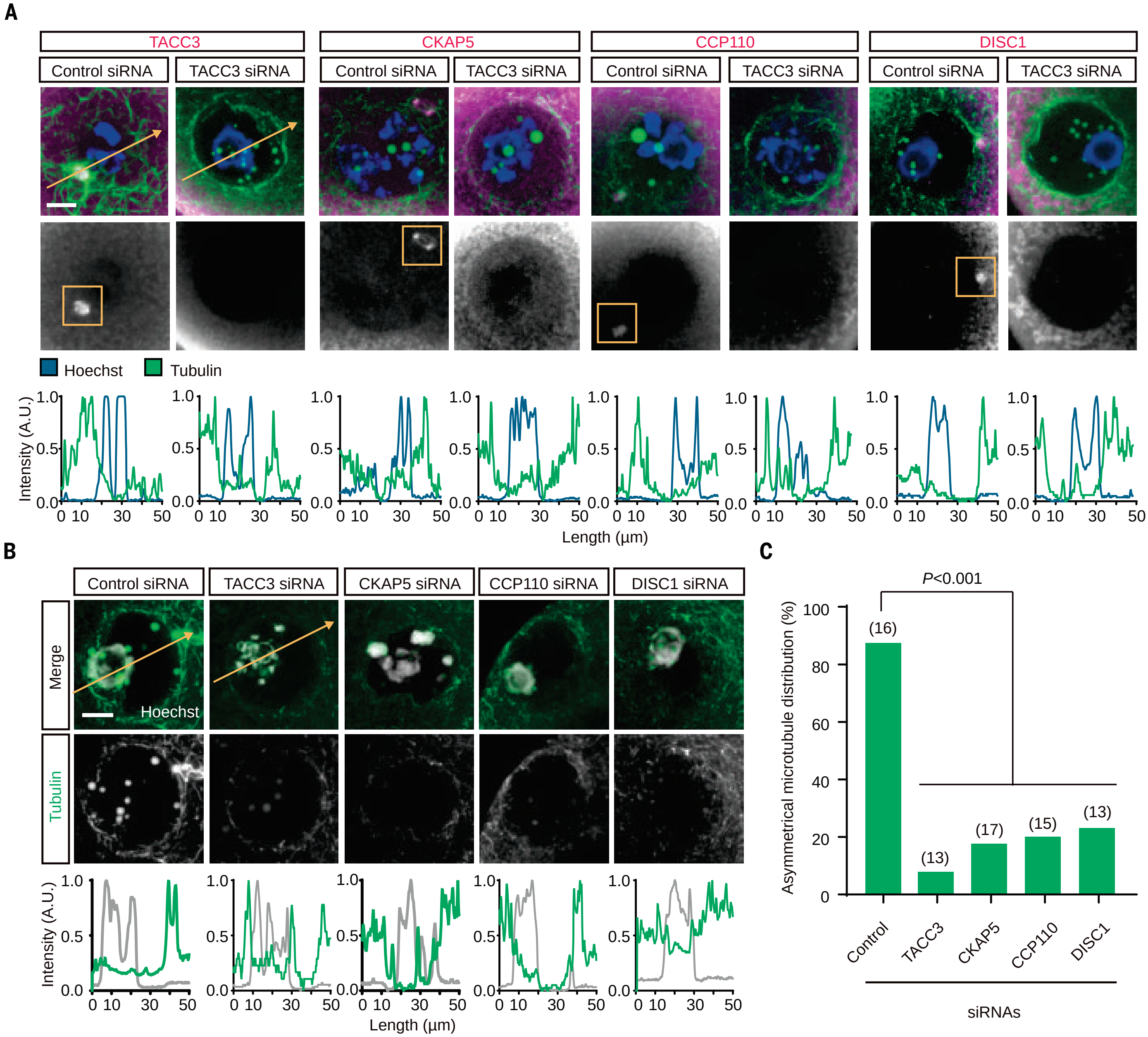


Fig. 3. The huoMTOC is required for microtubule nucleation in human oocytes. (**A**) Immunofluorescence images of human GV oocytes injected with TACC3 siRNAs. Green, microtubules (tubulin); magenta, huoMTOC (yellow squares); blue, chromatin. Scale bar, 10 μm. The measurements of tubulin intensity along the direction of the yellow arrows are shown at the bottom. Green, tubulin; blue, chromosomes. (**B**) Immunofluorescence images of human GV oocytes injected with specific siRNAs. Green, microtubules (tubulin); gray, chromosomes (Hoechst). The yellow arrows indicate the direction of the tubulin intensity measurements shown at the bottom. Scale bar, 10 μm. (**C**) The percentages of human oocytes with asymmetrical microtubule distribution in (B). $P < 0.001$, Fisher's exact test. Data were from three independent experiments. The number of oocytes analyzed is specified in parentheses.

meiotic spindle was completely disrupted (Fig. 5C). We also determined the stability of the huoMTOC and microtubule distribution in the patients' GV oocytes. Compared to normal human GV oocytes, the huoMTOC was missing, and the asymmetrical distribution of microtubules around the nucleus was impaired in the patients' GV oocytes (Fig. 5D).

In addition, supplementing the wild-type *TACC3* mRNA successfully rescued the phenotype of spindle disruption resulting from TACC3 depletion, whereas supplementing the mutant *TACC3* mRNAs could not rescue the phenotype (Fig. 5, E and F), suggesting that the mutations had loss-of-function effects on TACC3. Thus, TACC3 deficiency caused female infertility and oocyte maturation arrest by disrupting the integrity of the huoMTOC. These results suggest that disruption of the huoMTOC impaired the nucleation of microtubules in the GV oocytes of patients, further highlighting the critical role of the huoMTOC in regulating microtubule nucleation and the initiation of acentrosomal spindle assembly.

Discussion

Here, we report a structure that we named the huoMTOC, which serves as a major site of microtubule nucleation and is required for spindle assembly in human oocytes. A single huoMTOC is formed near the cortex of human GV oocytes at the time of meiosis resumption, and it migrates to the nuclear envelope before

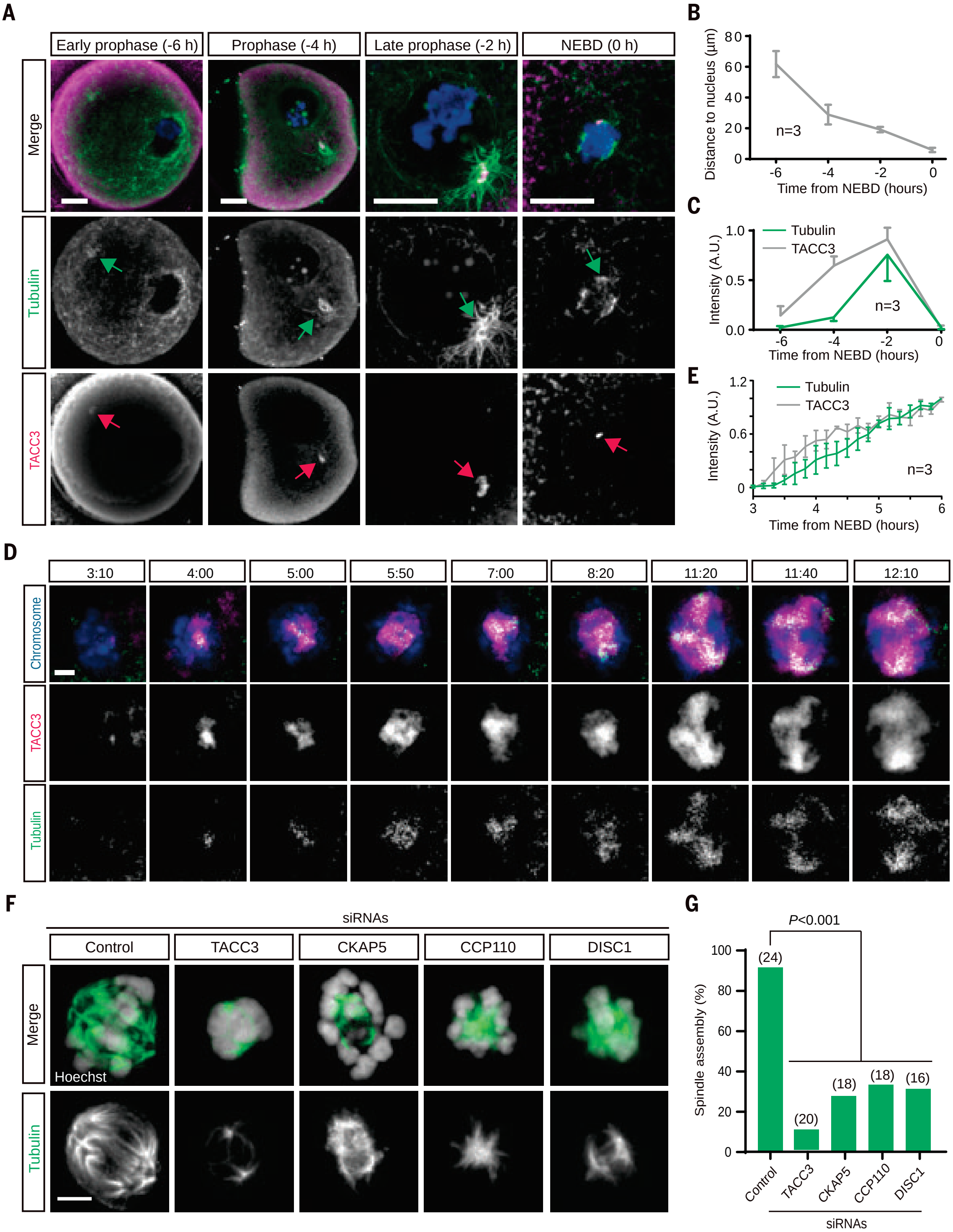


Fig. 4. The huoMTOC is recruited from the cortex to kinetochores for spindle microtubule polymerization in human oocytes. (**A**) Immunofluorescence images of human GV oocytes at NEBD and at 2 hours (late prophase), 4 hours (prophase), and 6 hours (early prophase) before NEBD. Green, microtubules (tubulin); blue, chromatin (Hoechst); magenta, huoMTOC (TACC3). Green arrows indicate the microtubule cluster, and magenta arrows indicate the huoMTOC. Scale bar, 20 μm. (**B**) The distance between the huoMTOC and the nucleus was measured in (A). Time is after NEBD. (**C**) The intensity of the microtubule cluster (tubulin) and the huoMTOC (TACC3) was measured at different time points in (A). Time is after NEBD. (**D**) Representative time-lapse images showing the relationship between the huoMTOC and microtubule nucleation in live human oocytes. Green, microtubules (SiR-tubulin); blue, chromosomes (H2B-mClover); magenta, huoMTOC (mScarlet-TACC3). Scale bar, 5 μm. (**E**) The intensity of microtubules (SiR-tubulin) and huoMTOC (mClover-TACC3) close to chromosomes was measured in (D). Time is after NEBD. The number of oocytes analyzed in experiments is indicated. The mean and standard error were calculated on the basis of two independent experiments. Error bars are standard deviations. (**F**) Immunofluorescence images of human MI spindles from control and TACC3, CKAP5, CCP110, and DISC1 siRNA-injected human oocytes. Green, microtubules (tubulin); gray, chromosomes (Hoechst). Scale bar, 5 μm. (**G**) The spindle assembly percentage measured and collected from (F). The number of oocytes analyzed in three independent experiments is indicated. P < 0.001, Fisher's exact test.

NEBD. After NEBD, the huoMTOC becomes fragmented and is recruited to chromo- somes and kinetochores for spindle microtubule nucleation (Fig. 6). With the microtubule polymerization, the huoMTOC proteins are also recruited to the spindle microtubules (Fig. 6 and fig. S9C). Ablation of the huoMTOC results in microtubule loss and defective spindle assembly. In addition, we demonstrated that TACC3, CCP110, CKAP5, and DISC1 are essential components of the huoMTOC and that mutations in *TACC3* cause clinical oocyte maturation arrest and female infertility.

Distinct aMTOCs have been identified and investigated in mouse oocytes as microtubule

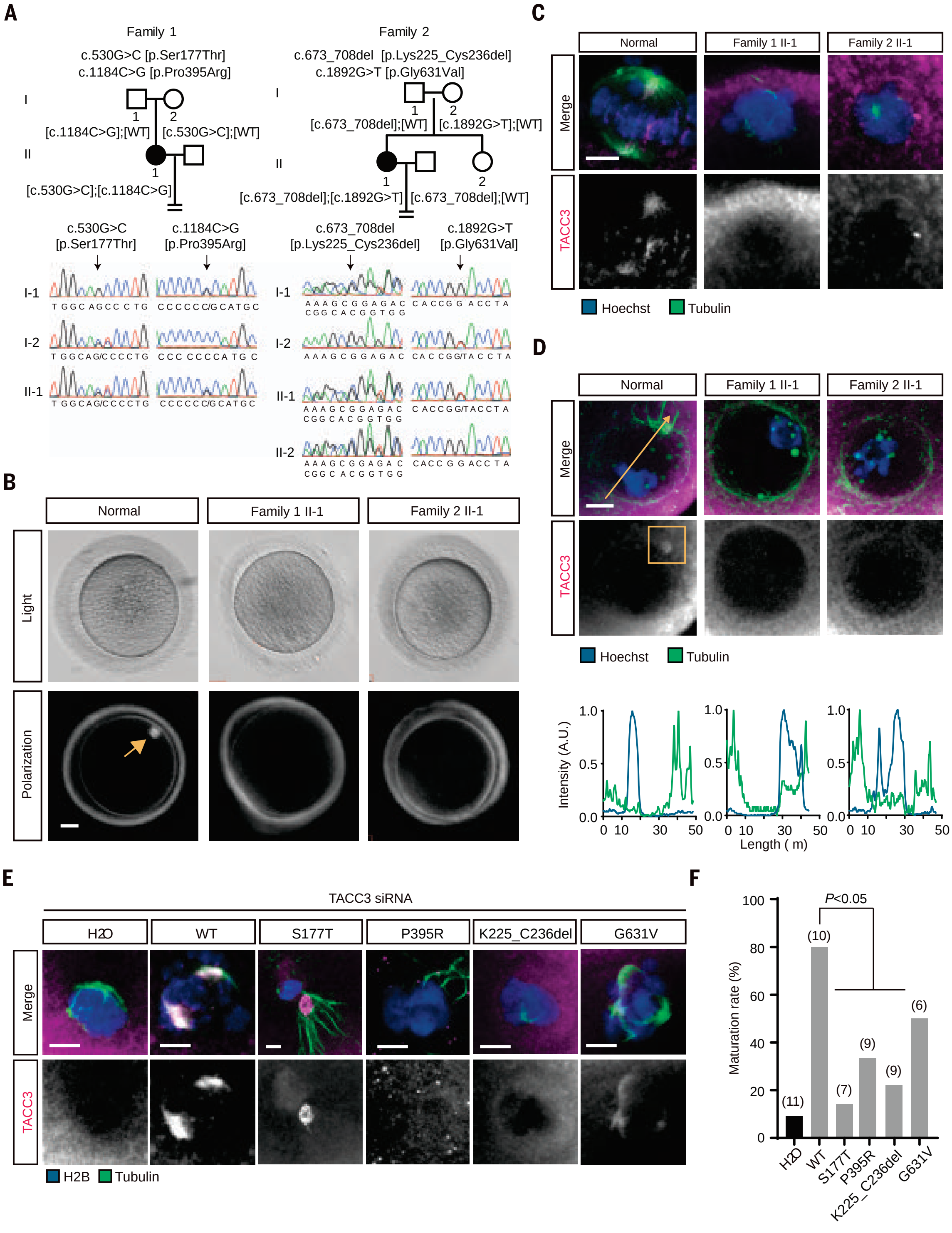


Fig. 5. Disruption of the huoMTOC in human oocytes impairs microtubule nucleation and spindle assembly. (**A**) Pedigrees of the two families with *TACC3* mutations with Sanger sequencing confirmation. Squares denote male family members, circles denote female members, black solid circles denote probands, and the equal sign denotes infertility. (**B**) Human oocytes from donors (normal) and patients (family 1 II-1, family 2 II-1) were examined by light and polarization microscopy. The arrow indicates an MI spindle. (**C**) Immunofluorescence images of human MI oocytes from donors (normal) and patients (family 1 II-1, family 2 II-1). Green, microtubules (tubulin); blue, chromosomes (Hoechst); magenta, TACC3. Scale bar, 5 μm. (**D**) Immunofluorescence images of human GV oocytes from donors (normal) and patients (family1 II-1, family2 II-1). Green, microtubules (tubulin); blue, chromosomes (Hoechst); magenta, TACC3. Scale bar, 10 μm. The yellow square shows the huoMTOC in normal human oocyte. The microtubule distribution was measured as previously described. (**E**) Immunofluorescence images of human MI oocytes injected with *TACC3* siRNAs and wild type or patient-derived mutant mRNAs. Green, microtubules (tubulin); blue, chromosomes (Hoechst), magenta, FLAG-TACC3. Scale bar, 5 μm. (**F**) The percentage of human oocyte maturation measured in (E). (Fisher's exact test, $P < 0.05$). The number of oocytes analyzed is specified in parentheses.

nucleators for meiotic spindle assembly (*15*). However, the primary spindle microtubule nucleator of human oocytes has remained unknown. The following factors might be reasons why the huoMTOC was not identified in previous investigations: (i) Unlike mouse aMTOCs, the huoMTOC is not concentrated on the spindle poles in human MI and MII oocytes (*22*). (ii) The classic aMTOC marker pericentrin is not among the components of the huoMTOC and therefore cannot label the structure. (iii) Only a single huoMTOC is formed in late prophase, and it is fragmented immediately after NEBD, making it difficult to capture in human oocytes by immunofluorescence.

In previous investigations, spindle assembly of human oocytes was reported to be mediated by chromosomes and dependent on

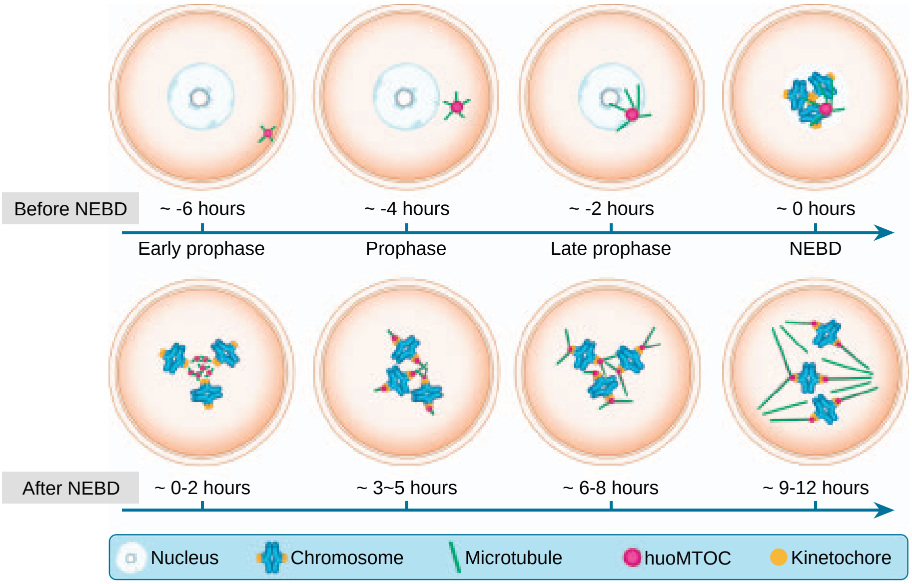


Fig. 6. Mechanistic model for huoMTOC migration and microtubule nucleation in human oocytes. The huoMTOC (magenta) assembles near the cortex of GV oocytes and migrates to the nuclear envelope before NEBD. The huoMTOC expands and the surrounding microtubules (green) keep growing until NEBD. After NEBD, the huoMTOC fragments, and the surrounding microtubules are disassembled. The fragmented huoMTOC is then recruited to kinetochores and initiates microtubule nucleation for meiotic spindle assembly.

RanGTP (*22*). It has been demonstrated that RanGTP inhibition impairs spindle assembly in human oocytes (*22*). In mouse or *Drosophila* oocytes, spindles have defects but still assemble if RanGTP is inhibited, suggesting that the RanGTP pathway is not essential for meiotic spindle assembly in oocytes of these two species (*35*, *36*). In the mouse oocytes lacking centrosomes, the processes of spindle assembly and spindle microtubule nucleation are mainly achieved by aMTOCs and facilitated by liquid-like meiotic spindle domain (*21*, *37*). It has been argued for a long time that, unlike mouse oocytes, human oocytes lack prominent aMTOCs (*2*, *22*, *38*). However, in the present study, the spindle microtubule in human oocytes was observed to be primarily polymerized from the kinetochores by an aMTOC-like structure that we named huoMTOC. In the *Drosophila* oocytes that also lack prominent aMTOCs, the spindle assembly is dominated by the chromosomes, which recruit and/or nucleate the microtubules (*39*). The spindle microtubules are organized around the chromosomes into fibers of two types, the interpolar and kinetochore microtubules, to assemble the meiotic spindle of *Drosophila* (*39*). Although the spindle assembly is directed by chromosomes in both human and *Drosophila* oocytes, the mechanisms involved could be different. The meiotic spindle assembly in *Drosophila* oocytes requires the chromosomal passenger complex (*16*), but that in human oocytes requires the huoMTOC.

In our study, the localization of 86 centrosome and spindle-related proteins was performed in human MI oocytes by systematic immunofluorescent staining. We identified the components (TACC3, CKAP5, CCP110, and DISC1) of the huoMTOC on kinetochores in MI oocytes and then verified their localization in GV oocytes using high-resolution imaging in live human oocytes. Emerging evidence suggests that the human TACC3 plays an important role in microtubule growth and spindle assembly during mitosis (*30*, *40*–*42*). The microtubule nucleation is blocked in ovarian cancer cells when TACC3 expression is affected, suggesting that TACC3 is required for centrosome-involved microtubule nucleation (*43*). TACC3 depletion impairs the γ-tubulin ring complex assembly (*44*) and centrosome integrity (*45*) of human somatic cells. CKAP5 was previously identified as a microtubule nucleation factor in vitro (*32*, *33*, *46*). CKAP5 was observed functioning synergistically with the γ-tubulin ring complex for de novo microtubule nucleation (*32*) and catalyzing numerous rounds of tubulin subunit addition at the microtubule plus-end for spindle microtubule assembly in vitro (*33*). In addition, CKAP5 was also implicated in the importin-regulated microtubule nucleation as a microtubule polymerase in vitro (*46*). In mitosis, TACC3 and CKAP5 are RanGTP-regulated spindle assembly factors for spindle microtubule nucleation and stabilization (*47*–*49*). Collectively, we inferred that TACC3 and CKAP5 may act as organizers of huoMTOC assembly and microtubule nucleation in human oocytes. The functions of CCP110 and DISC1 in microtubule polymerization have not yet been determined in either mitotic or meiotic cells.

Apart from the four huoMTOC components, we also identified 39 other proteins that showed spindle-related localization. We therefore cannot exclude the possibility that some of these proteins may also play a role in huoMTOC formation and function. The specific components of huoMTOC and the mechanism for its nucleation of microtubules are worth investigating in the future. In addition, according to our observations, the localization of most centrosome and microtubule-related proteins in the MI spindle of human oocytes is obviously different from that in mouse oocytes (*21*). Thus, future investigations on the functions of these proteins in human oocytes should shed more light on the mechanism of human oocyte spindle assembly.

In a recent study, a distinct mechanism of spindle pole organization was discovered in human oocytes, suggesting that loss of kinesin superfamily protein C1 (KIFC1) induces meiotic spindle instability (*23*). In addition, our previous investigations suggest that tubulin beta 8 class VIII (TUBB8) is the main isotype of spindle β-tubulin in human oocytes but is not found in mice or other nonprimate species (*50*). These findings suggest that a distinct mechanism for the initiation of microtubule nucleation and spindle assembly has evolved in human oocytes, which may contribute to a series of physiological characteristics including increasing chromosome segregation errors, high spindle instability, and aneuploidy in human oocytes.

In the clinic, many infertility patients experience recurrent failure of IVF or ICSI attempts owing to oocyte maturation arrest. However, the genetic factors involved remain unknown for most patients. In this study, we found that mutations in *TACC3* cause defects in spindle assembly by disrupting the structure of the huoMTOC, which leads to oocyte maturation arrest in patients. It is worth performing mutational screening for both known and potential genes that might participate in the maintenance of huoMTOC integrity. The results of such screening will help in precision diagnosis for these patients and will provide therapeutic targets for future clinical treatments. Our findings provide not only insights into the physiological mechanism of microtubule nucleation and spindle assembly in human oocytes but also improve our understanding of pathophysiological mechanisms of human oocyte maturation arrest.

Materials and methods

Human oocyte collection and culture

Human GV oocytes were donated by patients undergoing ICSI as part of their assisted reproduction treatment at Shanghai Ji Ai Genetics and IVF Institute affiliated with the Obstetrics and Gynecology Hospital of Fudan University,

Center for Reproductive Medicine and Fertility Preservation program affiliated with International Peace Maternity and Child Health Hospital of Shanghai Jiao Tong University. Only immature oocytes that were unable to be used for assisted reproduction treatment were collected for this research, and the use of these human GV oocytes was clearly explained to the patient donors. All these female donors were receiving ICSI treatment because of male factor-induced infertility. The age range of the donors was 25 to 38 and their mean age was 34.3. The BMI range of these donors was 20 to 22. The collected human oocytes were mixed thoroughly and distributed randomly to the control and experimental groups. Only morphologically normal human GV oocytes were used in this investigation. G-MOPS medium (Vitrolife) with milrinone (2 μM, HY-14252, MedChemExpress) was used to maintain human oocyte prophase arrest. The human GV oocytes were matured in Multipurpose Handling Medium-Complete (MHM-C) (FUJIFILM Irvine Scientific) or G-MOPS medium at 37°C on a heating block.

Animals and oocyte culture

The porcine ovaries were collected from local slaughterhouses and transported in warm 0.9% NaCl. Porcine GV oocytes were collected in TCM199 medium at 39°C, and milrinone (2 μM) was added to the medium to maintain oocyte prophase arrest. Only fully grown oocytes were used in our experiments. For maturation, GV oocytes were washed free from milrinone and cultured in fresh TCM199 medium at 39°C and 5% CO_2. Porcine GV and MI oocytes were fixed for immunofluorescence at 6 or 15 hours after oocyte isolation, respectively.

Female C57Bl/6 mice (3- to 4-week-old) were purchased from Beijing Vital River Laboratory Animal Technology Co., Ltd. (Beijing, China). All mice were used in accordance with institutional guidelines, and the experiments were approved by the Animal Care and Use Committee of Fudan University, China. Mouse GV oocytes were released from the ovaries at 44 to 52 hours after injection of 10 IU pregnant mare serum gonadotropin. For maturation, denuded GV oocytes were cultured in fresh M2 medium at 37°C on a heat block. The pH value of culture medium was between 7.2 and 7.4 (S210, Mettler Toledo). Mouse GV and MI oocytes were fixed for immunofluorescence at 0.5 or 6 hours after in vitro maturation initiation, respectively.

Inhibitor treatment

For microtubule depolymerization, nocodazole (10 μM, HY-13520, MedChemExpress) was added to human oocytes at ~10 hours after NEBD and 1 hour before time-lapse imaging. The drug was dissolved in dimethyl sulfoxide (Sigma-Aldrich) and used at a concentration of 0.1% in G-MOPS.

Immunofluorescence microscopy

Human GV oocytes were fixed for immunofluorescence at 2, 4, or 6 hours after milrinone washout. Human MI oocytes were fixed at 5 to 12 hours after NEBD (prometaphase or metaphase). Human oocytes were fixed for 30 min in phosphate-buffered saline (PBS) containing 2% formaldehyde and 0.1% Triton-X at 37°C on a heat block and were then permeabilized in PBS containing 0.5% Triton-X (PBT) at 4°C overnight. Oocytes were extensively washed with PBS between stages and then blocked at room temperature in a blocking buffer of 3% bovine serum albumin (BSA) in 0.3% PBT. All antibody incubations were performed in blocking buffer at 4°C overnight for primary antibodies and at 37°C for 1 hour for secondary antibodies. Primary antibodies were anti-centromere antibody (HCT-0100; ImmunoVision; 1:500), anti-AKAP450 antibody (611518; BD Biosciences; 1:50), anti-ASPM antibody (26223-1-AP; Proteintech; 1:50), anti-AURKA antibody (NBP2-50041; Novus Biological; 1:50), anti-AURKA antibody (10297-1-AP; Proteintech; 1:20), anti-AURKB antibody (ab45145; Abcam; 1:50), anti-BBS4 antibody (12766-1-AP; Proteintech; 1:50), anti-beta tubulin antibody (ab204686; Abcam; 1:50), anti-beta tubulin antibody (ab11309; Abcam; 1:100), anti-BUGZ antibody (A20177; ABclonal; 1:20), anti-CAMSAP2 antibody (17880-1-AP; Proteintech; 1:50), anti-CCP110 antibody (12780-1-AP; Proteintech; 1:50), anti-CCP110 antibody (PA5-58775; Thermo Fisher Scientific; 1:100), anti-CDK5RAP2 antibody (06-1398; Merck; 1:100), anti-CENPJ antibody (11517-1-AP; Proteintech; 1:50), anti-CEP57 antibody (24957-1-AP; Proteintech; 1:50), anti-CEP63 antibody (16268-1-AP; Proteintech; 1:50), anti-CEP72 antibody (19928-1-AP; Proteintech; 1:50), anti-CEP120 antibody (PA5-55985; Thermo Fisher Scientific; 1:100), anti-CEP135 antibody (24428-1-AP; Proteintech; 1:50), anti-CEP152 antibody (21815-1-AP; Proteintech; 1:50), anti-CEP164 antibody (22227-1-AP; Proteintech; 1:50), anti-CEP170 antibody (27325-1-AP; Proteintech; 1:50), anti-CEP192 antibody (18832-1-AP; Proteintech; 1:50), anti-CEP250 antibody (14498-1-AP; Proteintech; 1:50), anti-CEP290 antibody (22490-1-AP; Proteintech; 1:50), anti-CETN2 antibody (A5397; ABclonal; 1:20), anti-CETN3 antibody (A8111; ABclonal; 1:50), anti-CKAP5 antibody (PA5-59150; Thermo Fisher Scientific; 1:100), anti-CKAP5 antibody (CL488-67631; Proteintech; 1:50), anti-CLIP1 antibody (23839-1-AP; Proteintech; 1:50), anti-CLTC antibody (610500; BD Biosciences; 1:50), anti-CNTROB antibody (26880-1-AP; Proteintech; 1:50), anti-DCTN1 antibody (55182-1-AP; Proteintech; 1:50), anti-DCTN2 antibody (A2200; ABclonal; 1:50), anti-DYNC1H1 antibody (12345-1-AP; Proteintech; 1:50), anti-DISC1 antibody (A4678; ABclonal; 1:50), anti-DLGAP5 antibody (A13575; ABclonal; 1:50), anti-DYNLT1 antibody (11954-1-AP; Proteintech; 1:50), anti-GTSE1 antibody (A302-425A; Bethyl Laboratories; 1:50), anti-HAUS4 antibody (20104-1-AP; Proteintech; 1:50), anti-HAUS6 antibody (A4797; ABclonal; 1:50), anti-HAUS8 antibody (PA5-21331; Thermo Fisher Scientific; 1:100), anti-HMMR antibody (15820-1-AP; Proteintech; 1:50), anti-HOOK2 antibody (ab133691; Abcam; 1:50), anti-HOOK3 antibody (A15536; ABclonal; 1:50), anti-HOOK3 antibody (15457-1-AP; Proteintech; 1:50), anti-KANSL3 antibody (HPA035018; Merck; 1:100), anti-KIF11 antibody (HPA010568; Merck; 1:100), anti-KIF20A antibody (15911-1-AP; Proteintech; 1:50), anti-KIF20A antibody (CL594-67190; Proteintech; 1:50), anti-KIF22 antibody (A19881; ABclonal; 1:50), anti-KIF2A antibody (13105-1-AP; Proteintech; 1:50), anti-KIF2B antibody (A6480; ABclonal; 1:20), anti-KIFC1 antibody (A3304; ABclonal; 1:20), anti-KIZ antibody (21177-1-AP; Proteintech; 1:50), anti-LIS1 antibody (H00005048-M03; Abnova; 1:50), anti-LMNB antibody (A1910; ABclonal; 1:20), anti-LRRC45 antibody (PA5-54777; Thermo Fisher Scientific; 1:100), anti-MCRS1 antibody (HPA039057; Merck; 1:100), anti-MYO10 antibody (sc-23137; SantaCruz Biotechnology; 1:50), anti-NDE1 antibody (10233-1-AP; Proteintech; 1:50), anti-NDEL1 antibody (H00081565-D01P; Abnova; 1:50), anti-NEDD1 antibody (13993-1-AP; Proteintech; 1:50), anti-NEK2 antibody (14233-1-AP; Proteintech; 1:50), anti-NIN antibody (A8215; ABclonal; 1:50), anti-NUMA1 antibody (A0527; ABclonal; 1:50), anti-NUP107 antibody (A13110; ABclonal; 1:50), anti-NUP160 antibody (PRS4707; Merck; 1:100), anti-NUP62 antibody (13916-1-AP; Proteintech; 1:50), anti- NUP85 antibody (19370-1-AP; Proteintech; 1:50), anti-NUSAP1 antibody (12024-1-AP; Proteintech; 1:50), anti-ODF2 antibody (12058-1-AP; Proteintech; 1:50), anti-PCNT antibody (611815; BD Biosciences; 1:200), anti-PCM1 antibody (19856-1-AP; Proteintech; 1:50), anti-PLK1 antibody (A2548; ABclonal; 1:50), anti-PLK3 antibody (10977-1-AP; Proteintech; 1:50), anti-PLK4 antibody (12952-1-AP; Proteintech; 1:50), anti-PRC1 antibody (15617-1-AP; Proteintech; 1:50), anti-RAE1 antibody (20491-1-AP; Proteintech; 1:50), anti-RAN antibody (10469-1-AP; Proteintech; 1:50), anti-SASS6 antibody (21377-1-AP; Proteintech; 1:50), anti-SKAP2 antibody (12926-1-AP; Proteintech; 1:50), anti-SNF2H antibody (A2000; ABclonal; 1:50), anti-SPDL1 antibody (PA5-99285; Thermo Fisher Scientific; 1:100), anti-SSX2IP antibody (13694-1-AP; Proteintech; 1:50), anti-TACC3 antibody (A18641; ABclonal; 1:50), anti-TACC3 antibody (ab134154; Abcam; 1:100), anti-TOP2A antibody (20233-1-AP; Proteintech; 1:50), anti-TPX2 antibody (A18327; ABclonal; 1:20), anti-TUBG1 antibody (A9657; ABclonal; 1:20). Secondary antibodies were Alexa Fluor 647-conjugated anti-rabbit IgG (4414S; Cell Signaling; 1:200), Alexa Fluor 647-conjugated

anti-mouse IgG (4410S; Cell Signaling; 1:200), Cy3-conjugated anti-rabbit IgG (AS008; ABclonal; 1:500), and Atto 488-conjugated anti-human IgG (52526; Merck; 1:500). Chromatin was briefly counterstained with Hoechst 33342 (20 μg·ml^{-1}, HY-15559, MedChemExpres) before imaging. The samples were in PBS and imaged with an LSM 880 confocal laser scanning microscope (ZEISS) with a 63×/1.4 NA Plan Apochromat oil immersion lens at room temperature.

RNA interference

The siRNAs were provided by GenePharma or Tsingke Biotechnology. The sequences of siRNA for TACC3 down-regulation were 5′-GGU UCG AAG AGG UUG UGU A-3′ and 5′-GCA UGC ACG GUG CAA AUG A-3′. The siRNA sequences against CKAP5 were 5′-GGA AAT AGC TGT TCA CAT A-3′ and 5′-GGC CAA AGC TCC AGG ATT A-3′. The siRNA sequences targeting CCP110 were 5′-CAC UCU ACU GCA GCA AAG C-3′ and 5′-AUG UUC UUC UCC AAG GUG C-3′. The siRNA sequence targeting DISC1 was 5′-GGA UUU GAG AAU AGU UUC A-3′. The negative control siRNAs were provided by the same company. To increase the efficiency of RNA inhibition, mixed siRNAs were microinjected into human GV oocytes. The final concentration of siRNAs was 40 μM.

Microinjection of human oocytes

Human GV oocytes were microinjected in G-MOPS with milrinone (2 μM) on the stage of an inverted microscope (Leica) with micromanipulators (Eppendorf). A 0.1 to 0.3% volume of mRNA was injected using a timed pulse, and the final concentration of mRNA was 1 μg/μl. The injected GV oocytes were arrested in prophase for 2–4 hours for mRNA expression.

mRNA synthesis

mRNA was transcribed in vitro from purified linear double-stranded DNA templates. mMessage T7 or T3 RNA polymerase kits (New England Biolabs) were used for the in vitro transcription reaction. The constructs of mClover3-CENPB, mScarlet-CENPB, mClover3-TACC3, mScarlet-TACC3, mClover3-CKAP5, mClover3-CCP110, and mClover3-DISC1 were made and used for mRNA production.

Live cell imaging

For high-resolution time-lapse imaging, time points were acquired at 10-min intervals using an LSM 880 confocal laser scanning microscope (ZEISS) fitted with sensitive detectors, an environmental chamber set to 37°C, and a long-distance 40×/1.1 NA C-Apochromat water immersion lens. A volume of 30 μm by 30 μm by 15 μm centered around the chromosomes was typically imaged. The chromosomes were tracked automatically by MyPiC on LSM880. Microtubules in live human oocytes were visualized by SiR-tubulin (1 μM) staining in G-MOPS. To reduce background noise, some images were passed through a Gaussian filter of 2 sigma in Fiji (NIH).

Fluorescent intensity measurement

The pattern of microtubule distribution was defined by the intensity of tubulin around the nuclear envelope. The direction of intensity measurements is shown in the figures. The intensity was measured by Image J (NIH). The integrated intensity of kinetochore foci was measured with the Foci_Picker3D plugin in Image J. The same threshold was applied to each focus within an oocyte.

Laser ablation

To examine the role of huoMTOC in microtubule polymerization of human oocytes, the huoMTOC was directly disrupted by laser in live human GV oocytes. The human GV oocytes expressing mClover3-TACC3 were rotated with an unbroken microinjection pipette to obtain huoMTOC. The square regions of interest were marked and photobleached using a 488-nm laser line at the maximum power. The laser ablation was performed at 37°C.

Cell culture and transfection

HEK293T cells were obtained from the Cell Bank of Shanghai Institute for Biological Sciences, the Chinese Academy of Sciences (Shanghai, China). Cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin (Gibco, Waltham, MA, USA) in an atmosphere of 5% CO_2 at 37°C to between 70 and 80% confluence. Plasmids were transfected into HEK293T cells using the PolyJet In Vitro DNA Transfection Reagent (SignaGen) according to the manufacturer's instructions.

Immunoblots and immunoprecipitation

HEK293T cells were harvested after transfection for 36 hours and washed with PBS. Cells were lysed in radioimmunoprecipitation assay lysis buffer (Shanghai Wei AO Biological Technology, Shanghai, China) with 1% protease inhibitor cocktail (Bimake, Houston, TX, USA). After quantification with the bicinchoninic acid assay (Shanghai Biocolor BioScience & Technology Co.), the supernatant was subjected to immunoprecipitation with affinity beads (Sigma). After incubation at 4°C for 4 hours, the beads were washed with lysis buffer four times.

The bead-bound proteins were eluted using sodium dodecyl sulfate (SDS) sample buffer, resolved by SDS–polyacrylamide gel electrophoresis (SDS-PAGE), transferred to nitrocellulose membranes (Pall Corporation), and probed with rabbit anti-FLAG (1:3000 dilution; Cell Signaling Technology) or mouse anti-vinculin (1:5000 dilution; Sigma-Aldrich) antibodies. The secondary antibodies were goat anti-rabbit immunoglobulin G (IgG) (1:5000 dilution; Abmart) or goat anti-mouse IgG (1:5000 dilution; Abmart) conjugated to horseradish peroxidase.

Clinical samples

A cohort of 1394 infertile female patients with oocyte maturation arrest recruited from the Ninth Hospital affiliated with Shanghai Jiao Tong University and Shanghai Ji Ai Genetics and IVF Institute affiliated with the Obstetrics and Gynecology Hospital of Fudan University participated in this study. Written informed consent was provided by patients. The recruitment of patients was performed as follows: (i) female patients were younger than 45 years old, failing to conceive after 1 year (or longer) of regular unprotected sex; (ii) had undergone ≥2 failed attempts of IVF/ICSI, characterized by oocyte maturation arrest; (iii) female patients with other known causes of infertility, including male factors, chromosome anomalies, radiotherapy, or chemotherapy, were excluded. Peripheral blood samples were taken for DNA extraction.

The GV and MI oocytes from two patients with compound heterozygous mutations in *TACC3* were obtained as part of their assisted reproduction treatment at the Shanghai Ninth Hospital affiliated to Shanghai Jiao Tong University.

This study was approved by the Ethics Committee of the Medical College of Fudan University and the Reproductive Study Ethics Committees of the hospitals.

Genetic studies

Genomic DNA was extracted from peripheral blood using the QIAamp DNA Blood Mini Kit (Qiagen). Whole-exome capture was performed using the SeqCap EZ Exome Kit (Roche), and sequencing was performed on the Illumina NovaSeq 6000 platform (Illumina). Sequencing analysis was compared with the human reference sequence (NCBI Genome build GRCh37). Mutations were annotated with GRCh37 and the dbSNP (version 138) and gnomAD along with our in-house exome database (data deposited in public database, accession numbers GVM000402 and GVM000394).

Statistical analysis

The investigators were not blinded during experiments and outcome assessment. The experiments were not randomized. The statistical methods were not used to determine sample size. Sample means were compared with either Student's *t* test or one-way analysis of variance (ANOVA) with a post-hoc test as stated (two-sided). Dichotomous data were compared using Fisher's exact test (two-tailed). All data are from at least two independent experiments. All tests were performed using GraphPad Prism 7 (GraphPad Software).

REFERENCES AND NOTES

1. S. L. Prosser, L. Pelletier, Mitotic spindle assembly in animal cells: A fine balancing act. *Nat. Rev. Mol. Cell Biol.* **18**, 187–201 (2017). doi: 10.1038/nrm.2016.162; pmid: 28174430
2. B. Mogessie, K. Scheffler, M. Schuh, Assembly and positioning of the oocyte meiotic spindle. *Annu. Rev. Cell Dev. Biol.* **34**, 381–403 (2018). doi: 10.1146/annurev-cellbio-100616-060553; pmid: 30028643
3. S. Doxsey, Re-evaluating centrosome function. *Nat. Rev. Mol. Cell Biol.* **2**, 688–698 (2001). doi: 10.1038/35089575; pmid: 11533726
4. S. Petry, Mechanisms of mitotic spindle assembly. *Annu. Rev. Biochem.* **85**, 659–683 (2016). doi: 10.1146/annurev-biochem-060815-014528; pmid: 27145846
5. A. Vertii, H. Hehnly, S. Doxsey, The centrosome, a multitalented renaissance organelle. *Cold Spring Harb. Perspect. Biol.* **8**, a025049 (2016). doi: 10.1101/cshperspect.a025049; pmid: 27908937
6. J. L. Badano, T. M. Teslovich, N. Katsanis, The centrosome in human genetic disease. *Nat. Rev. Genet.* **6**, 194–205 (2005). doi: 10.1038/nrg1557; pmid: 15738963
7. D. Szollosi, P. Calarco, R. P. Donahue, Absence of centrioles in the first and second meiotic spindles of mouse oocytes. *J. Cell Sci.* **11**, 521–541 (1972). doi: 10.1242/jcs.11.2.521; pmid: 5076360
8. A. T. Hertig, E. C. Adams, Studies on the human oocyte and its follicle. I. Ultrastructural and histochemical observations on the primordial follicle stage. *J. Cell Biol.* **34**, 647–675 (1967). doi: 10.1083/jcb.34.2.647; pmid: 4292010
9. T. Mikeladze-Dvali *et al*., Analysis of centriole elimination during *C. elegans* oogenesis. *Development* **139**, 1670–1679 (2012). doi: 10.1242/dev.075440; pmid: 22492357
10. A. Pimenta-Marques *et al*., A mechanism for the elimination of the female gamete centrosome in *Drosophila melanogaster*. *Science* **353**, aaf4866 (2016). doi: 10.1126/science.aaf4866; pmid: 27229142
11. G. Manandhar, H. Schatten, P. Sutovsky, Centrosome reduction during gametogenesis and its significance. *Biol. Reprod.* **72**, 2–13 (2005). doi: 10.1095/biolreprod.104.031245; pmid: 15385423
12. D. L. Gard, Microtubule organization during maturation of *Xenopus* oocytes: Assembly and rotation of the meiotic spindles. *Dev. Biol.* **151**, 516–530 (1992). doi: 10.1016/0012-1606(92)90190-R; pmid: 1601183
13. M. Łuksza, I. Queguigner, M. H. Verlhac, S. Brunet, Rebuilding MTOCs upon centriole loss during mouse oogenesis. *Dev. Biol.* **382**, 48–56 (2013). doi: 10.1016/j.ydbio.2013.07.029; pmid: 23954884
14. M. J. Carabatsos, C. M. H. Combelles, S. M. Messinger, D. F. Albertini, Sorting and reorganization of centrosomes during oocyte maturation in the mouse. *Microsc. Res. Tech.* **49**, 435–444 (2000). doi: 10.1002/(SICI)1097-0029(20000601)49:5<435::AID-JEMT5>3.0.CO;2-H; pmid: 10842370
15. M. Schuh, J. Ellenberg, Self-organization of MTOCs replaces centrosome function during acentrosomal spindle assembly in live mouse oocytes. *Cell* **130**, 484–498 (2007). doi: 10.1016/j.cell.2007.06.025; pmid: 17693257
16. S. J. Radford, A. L. Nguyen, K. Schindler, K. S. McKim, The chromosomal basis of meiotic acentrosomal spindle assembly and function in oocytes. *Chromosoma* **126**, 351–364 (2017). doi: 10.1007/s00412-016-0618-1; pmid: 27837282
17. I. D. Wolff, M. V. Tran, T. J. Mullen, A. M. Villeneuve, S. M. Wignall, Assembly of *Caenorhabditis elegans* acentrosomal spindles occurs without evident microtubule-organizing centers and requires microtubule sorting by KLP-18/kinesin-12 and MESP-1. *Mol. Biol. Cell* **27**, 3122–3131 (2016). doi: 10.1091/mbc.e16-05-0291; pmid: 27559133
18. D. Clift, M. Schuh, A three-step MTOC fragmentation mechanism facilitates bipolar spindle assembly in mouse oocytes. *Nat. Commun.* **6**, 7217 (2015). doi: 10.1038/ncomms8217; pmid: 26147444
19. T. Wu, S. I. R. Lane, S. L. Morgan, K. T. Jones, Spindle tubulin and MTOC asymmetries may explain meiotic drive in oocytes. *Nat. Commun.* **9**, 2952 (2018). doi: 10.1038/s41467-018-05338-7; pmid: 30054463
20. C. Gueth-Hallonet *et al*., γ-Tubulin is present in acentriolar MTOCs during early mouse development. *J. Cell Sci.* **105**, 157–166 (1993). doi: 10.1242/jcs.105.1.157; pmid: 8360270
21. C. So *et al*., A liquid-like spindle domain promotes acentrosomal spindle assembly in mammalian oocytes. *Science* **364**, eaat9557 (2019). doi: 10.1126/science.aat9557; pmid: 31249032
22. Z. Holubcová, M. Blayney, K. Elder, M. Schuh, Error-prone chromosome-mediated spindle assembly favors chromosome segregation defects in human oocytes. *Science* **348**, 1143–1147 (2015). doi: 10.1126/science.aaa9529; pmid: 26045437
23. C. So *et al*., Mechanism of spindle pole organization and instability in human oocytes. *Science* **375**, eabj3944 (2022). doi: 10.1126/science.abj3944; pmid: 35143306
24. T. Wu, S. I. R. Lane, S. L. Morgan, F. Tang, K. T. Jones, Loss of centromeric RNA activates the spindle assembly checkpoint in mammalian female meiosis I. *J. Cell Biol.* **220**, e202011153 (2021). doi: 10.1083/jcb.202011153; pmid: 34379093
25. J. R. A. Hutchins *et al*., Systematic analysis of human protein complexes identifies chromosome segregation proteins. *Science* **328**, 593–599 (2010). doi: 10.1126/science.1181348; pmid: 20360068
26. H. V. Goodson, E. M. Jonasson, Microtubules and microtubule-associated proteins. *Cold Spring Harb. Perspect. Biol.* **10**, a022608 (2018). doi: 10.1101/cshperspect.a022608; pmid: 29858272
27. Z. Chen, V. B. Indjeian, M. McManus, L. Wang, B. D. Dynlacht, CP110, a cell cycle-dependent CDK substrate, regulates centrosome duplication in human cells. *Dev. Cell* **3**, 339–350 (2002). doi: 10.1016/S1534-5807(02)00258-7; pmid: 12361598
28. F. Gergely *et al*., The TACC domain identifies a family of centrosomal proteins that can interact with microtubules. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14352–14357 (2000). doi: 10.1073/pnas.97.26.14352; pmid: 11121038
29. F. Gergely, V. M. Draviam, J. W. Raff, The ch-TOG/XMAP215 protein is essential for spindle pole organization in human somatic cells. *Genes Dev.* **17**, 336–341 (2003). doi: 10.1101/gad.245603; pmid: 12569123
30. C. H. Lin, C. K. Hu, H. M. Shih, Clathrin heavy chain mediates TACC3 targeting to mitotic spindles to ensure spindle stability. *J. Cell Biol.* **189**, 1097–1105 (2010). doi: 10.1083/jcb.200911120; pmid: 20566684
31. S. Petry, R. D. Vale, Microtubule nucleation at the centrosome and beyond. *Nat. Cell Biol.* **17**, 1089–1093 (2015). doi: 10.1038/ncb3220; pmid: 26316453
32. A. Thawani, R. S. Kadzik, S. Petry, XMAP215 is a microtubule nucleation factor that functions synergistically with the γ-tubulin ring complex. *Nat. Cell Biol.* **20**, 575–585 (2018). doi: 10.1038/s41556-018-0091-6; pmid: 29695792
33. G. J. Brouhard *et al*., XMAP215 is a processive microtubule polymerase. *Cell* **132**, 79–88 (2008). doi: 10.1016/j.cell.2007.11.043; pmid: 18191222
34. W. Fu *et al*., Self-assembly and sorting of acentrosomal microtubules by TACC3 facilitate kinetochore capture during the mitotic spindle assembly. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15295–15300 (2013). doi: 10.1073/pnas.1312382110; pmid: 24003142
35. J. Cesario, K. S. McKim, RanGTP is required for meiotic spindle organization and the initiation of embryonic development in Drosophila. *J. Cell Sci.* **124**, 3797–3810 (2011). doi: 10.1242/jcs.084855; pmid: 22100918
36. J. Dumont *et al*., A centriole- and RanGTP-independent spindle assembly pathway in meiosis I of vertebrate oocytes. *J. Cell Biol.* **176**, 295–305 (2007). doi: 10.1083/jcb.200605199; pmid: 17261848
37. C. So, S. Cheng, M. Schuh, Phase separation during germline development. *Trends Cell Biol.* **31**, 254–268 (2021). doi: 10.1016/j.tcb.2020.12.004; pmid: 33455855
38. A. Webster, M. Schuh, Mechanisms of aneuploidy in human eggs. *Trends Cell Biol.* **27**, 55–68 (2017). doi: 10.1016/j.tcb.2016.09.002; pmid: 27773484
39. S. Doubilet, K. S. McKim, Spindle assembly in the oocytes of mouse and *Drosophila*—similar solutions to a problem. *Chromosome Res.* **15**, 681–696 (2007). doi: 10.1007/s10577-007-1148-8; pmid: 17674154
40. I. Peset *et al*., Function and regulation of Maskin, a TACC family protein, in microtubule growth during mitosis. *J. Cell Biol.* **170**, 1057–1066 (2005). doi: 10.1083/jcb.200504037; pmid: 16172207
41. I. Peset, I. Vernos, The TACC proteins: TACC-ling microtubule dynamics and centrosome function. *Trends Cell Biol.* **18**, 379–388 (2008). doi: 10.1016/j.tcb.2008.06.005; pmid: 18656360
42. D. G. Booth, F. E. Hood, I. A. Prior, S. J. Royle, A TACC3/ch-TOG/clathrin complex stabilises kinetochore fibres by inter-microtubule bridging. *EMBO J.* **30**, 906–919 (2011). doi: 10.1038/emboj.2011.15; pmid: 21297582
43. R. Yao *et al*., A small compound targeting TACC3 revealed its different spatiotemporal contributions for spindle assembly in cancer cells. *Oncogene* **33**, 4242–4252 (2014). doi: 10.1038/onc.2013.382; pmid: 24077290
44. P. Singh, G. E. Thomas, K. K. Gireesh, T. K. Manna, TACC3 protein regulates microtubule nucleation by affecting γ-tubulin ring complexes. *J. Biol. Chem.* **289**, 31719–31735 (2014). doi: 10.1074/jbc.M114.575100; pmid: 25246530
45. T. V. Suhail, P. Singh, T. K. Manna, Suppression of centrosome protein TACC3 induces G1 arrest and cell death through activation of p38-p53-p21 stress signaling pathway. *Eur. J. Cell Biol.* **94**, 90–100 (2015). doi: 10.1016/j.ejcb.2014.12.001; pmid: 25613365
46. J. Roostalu, N. I. Cade, T. Surrey, Complementary activities of TPX2 and chTOG constitute an efficient importin-regulated microtubule nucleation module. *Nat. Cell Biol.* **17**, 1422–1434 (2015). doi: 10.1038/ncb3241; pmid: 26414402
47. P. R. Clarke, C. Zhang, Spatial and temporal coordination of mitosis by Ran GTPase. *Nat. Rev. Mol. Cell Biol.* **9**, 464–477 (2008). doi: 10.1038/nrm2410; pmid: 18478030
48. O. J. Gruss *et al*., Ran induces spindle assembly by reversing the inhibitory effect of importin α on TPX2 activity. *Cell* **104**, 83–93 (2001). doi: 10.1016/S0092-8674(01)00193-3; pmid: 11163242
49. P. Kalab, R. Heald, The RanGTP gradient – a GPS for the mitotic spindle. *J. Cell Sci.* **121**, 1577–1586 (2008). doi: 10.1242/jcs.005959; pmid: 18469014
50. R. Feng *et al*., Mutations in *TUBB8* and human oocyte meiotic arrest. *N. Engl. J. Med.* **374**, 223–232 (2016). doi: 10.1056/NEJMoa1510791; pmid: 26789871

ACKNOWLEDGMENTS

We thank all the patients and volunteers who participated in this study. We thank J. Dai from the Shanghai Academy of Agricultural Sciences for providing the porcine ovaries. We thank C. Zhang from Peking University for the gift of the human *TACC3* construct. We thank K. Qiao from the Cell Biological Imaging Core Facility of IMIB at Fudan University for assistance with imaging. We are grateful to X. Li (ZEISS) for technical support with confocal live-oocyte time-lapse imaging. **Funding:** This work was supported by the National Key Research and Development Program of China (2021YFC2700100), the Basic Science Center Program of the National Natural Science Foundation of China (82288102), the National Natural Science Foundation of China (81725006, 32130029, 82101737, 82171643, 81971450, and 81971382), the Shanghai Municipal Science and Technology Major Project (2017SHZDZX01), the Project of the Shanghai Municipal Science and Technology Commission (21XD1420300), the Capacity Building Planning Program for the Shanghai Women and Children's Health Service, and the collaborative innovation center project construction for Shanghai Women and Children's Health. **Ethics statement:** The studies involving human participants were reviewed and approved by the ethics committee of the Shanghai Ji Ai Genetics and IVF Institute (JIAI E2020-23 and JIAI E2022-05), the ethics committee of the International Peace Maternity and Child Health Hospital (GKLW 2021-18), and the ethics committee of Fudan University (FE21198 and 2020-012). The patients and participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by the animal welfare and ethics group of the Department of Experimental Animal Science of Fudan University (2017 1350 A265). **Author contributions:** T.W., J.D., J.F., and Y.K. contributed equally to this work. L.W., Q.S., and X.S. conceived of and designed the research study. T.W. designed the experiments and the methods for data analysis. T.W., J.D., Y.K., and H.G. performed the experiments. Y.L. and R.G. helped with the molecular work and genome sequencing. J.F., M.Z., and W.L. helped with human oocyte sample collection and manipulation. X.D. helped with human oocyte in vitro maturation. B.C. organized the medical records and analyzed the whole-exome data. L.W., Q.S., and T.W. drafted and revised the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data in the paper are present in the paper or the supplementary materials. In addition, the variation data reported in this paper

have been deposited in the Genome Variation Map of the National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences, under accession numbers GVM000402 (patients) and GVM000394 (controls). **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq7361
Figs. S1 to S11
Tables S1 and S2
MDAR Reproducibility Checklist
Movies S1 to S5

Submitted 27 April 2022; resubmitted 24 August 2022
Accepted 14 October 2022
10.1126/science.abq7361

RESEARCH ARTICLE SUMMARY

CANCER

Aberrant hyperexpression of the RNA binding protein FMRP in tumors mediates immune evasion

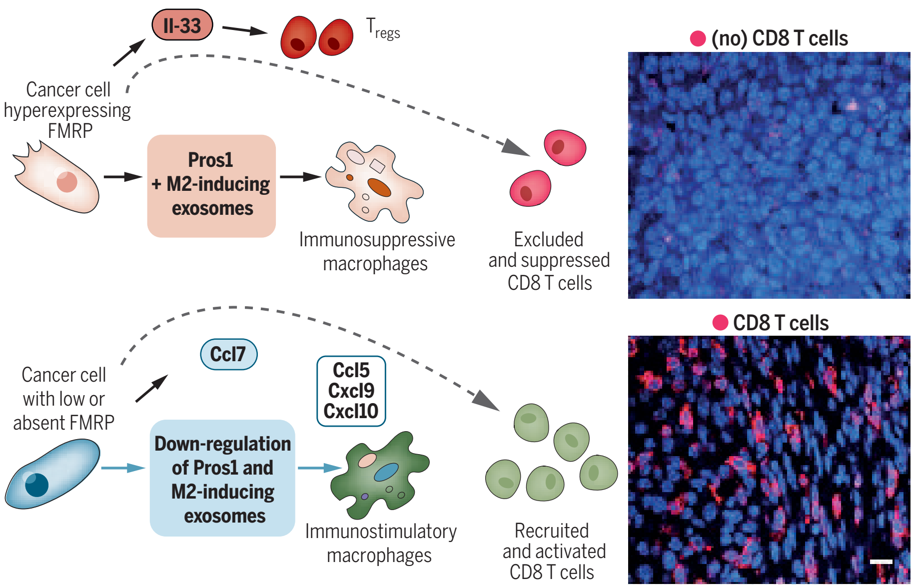
Qiqun Zeng†, Sadegh Saghafinia†, Agnieszka Chryplewicz†, Nadine Fournier, Lucine Christe, Yu-Qing Xie, Jeremy Guillot, Simge Yucel, Pumin Li, José A. Galván, Eva Karamitopoulou, Inti Zlobec, Dalya Ataca, Fleuriane Gallean, Peng Zhang, José Antonio Rodriguez-Calero, Mark Rubin, Mélanie Tichet, Krisztian Homicsko, Douglas Hanahan*

INTRODUCTION: Cancer biology and therapy have been transformed by knowledge about immunoregulatory mechanisms that govern adaptive immunity. Although some forms of treatment resistance are related to the intentionally transitory operations of the adaptive immune system, others reflect more subtle requirements to modulate the immune system in different contexts. In this work, we identified an immunoregulatory mechanism involving the neuronal RNA binding protein fragile X mental retardation protein (FMRP), which broadly regulates protein translation and mRNA stability and is aberrantly up-regulated in multiple forms of cancer.

RATIONALE: This study was motivated by reports that cancer cells naturally overexpressing FMRP, whose loss of expression in developing neurons causes cognitive defects, were invasive and metastatic. We investigated the expression of FMRP in human tumors, further assessed its tumor-promoting functions in mouse models of cancer, and evaluated its association with prognosis for human cancer patients.

RESULTS: When human tumor tissue microarrays were immunostained for expression of FMRP, a majority of tumors expressed FMRP, whereas cognate normal tissues did not. To investigate the functional significance of this broad up-regulation, the *FMR1* gene was ablated through CRISPR-Cas9 gene editing (FMRP-KO, where KO indicates knockout) in mouse cancer cell lines that were inoculated into both immunodeficient and syngeneic immunocompetent mice to establish tumors in parallel with wild-type (WT) FMRP-expressing cell lines. Mice bearing FMRP-KO tumors had similar survival compared with isogenic WT tumors in immunodeficient hosts, indicating that FMRP was not involved in stimulating tumor growth per se. By contrast, tumor growth was impaired and survival extended in immunocompetent hosts, implicating the adaptive immune system. Indeed, FMRP-expressing WT tumors were largely devoid of T cells, whereas FMRP-KO tumors were highly inflamed. Depletion of CD8 and CD4 T cells restored tumor growth and reduced survival, implicating FMRP in immune evasion in WT tumors. WT and FMRP-KO tumors were profiled by single-cell RNA sequencing, revealing marked differences in genome-wide transcription and abundance of cancer cells, macrophages, and T cells. To elucidate the effects of this multifaceted regulatory protein, we performed several functional perturbations, revealing that: FMRP-expressing cancer cells produce the chemokine interleukin-33 (IL-33), which induces regulatory T cells, as well as tumor-secreted protein S (PROS1) ligand and exosomes that elicit tumor-promoting (M2) macrophages. Both cell types are immunosuppressive, collectively contributing to the barrier against T cell attack. By contrast, FMRP-KO cancer cells down-regulate all three factors and up-regulate C-C motif chemokine ligand 7 (CCL7), which helps recruit and activate T cells. Additionally, immunostimulatory macrophages develop in this context that express three proinflammatory chemokines—CCL5, CXCL9, and CXCL10—which cooperate with CCL7 in recruiting T cells. Finally, neither *FMR1* mRNA nor FMRP protein levels were sufficient to predict outcomes in cohorts of cancer patients. Recognizing FMRP's function as an RNA binding protein that modulates mRNA stability and hence levels in transcriptome datasets, a gene signature reflecting FMRP's cancer regulatory activity (involving 156 genes) was developed by comparing FMRP-expressing versus FMRP-deficient cancer cells, both in culture and within tumors. Our FMRP cancer activity signature was prognostic for survival across multiple human cancers; anticorrelated with the intensity of T cell infiltration in different tumor types, consistent with FMRP's immunosuppressive effects; and was associated with comparatively poor responses to immune checkpoint inhibitors and immune-dependent chemotherapy in selected cohorts.

CONCLUSION: FMRP is revealed as a regulator of a network of genes and cells in the tumor microenvironment that contribute to the capability of tumors to evade immune destruction. ■



FMRP enables tumors to evade being attacked by the immune system. Up-regulated FMRP expression and activity are involved in an immunosuppressive program in cancer cells (upper left) that renders tumors impenetrable, creating so-called immune deserts (upper right). By contrast, its absence is associated with a reprogrammed tumor microenvironment that recruits and activates T lymphocytes, producing inflamed tumors, including CD8 T cells (red immunostain; lower right) with consequently beneficial immune destruction. T_{regs}, regulatory (immunosuppressive) T cells.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: douglas.hanahan@epfl.ch
Cite this article as Q. Zeng *et al.*, *Science* **378**, eabl7207 (2022). DOI: 10.1126/science.abl7207
†These authors contributed equally to this work.

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abl7207

RESEARCH ARTICLE

CANCER

Aberrant hyperexpression of the RNA binding protein FMRP in tumors mediates immune evasion

Qiqun Zeng[1,2]†, Sadegh Saghafinia[1,2]†, Agnieszka Chryplewicz[1,3]†, Nadine Fournier[4], Lucine Christe[5], Yu-Qing Xie[6], Jeremy Guillot[1,3], Simge Yucel[1,3], Pumin Li[1,3,7], José A. Galván[5], Eva Karamitopoulou[5], Inti Zlobec[5], Dalya Ataca[2], Fleuriane Gallean[2], Peng Zhang[8], José Antonio Rodriguez-Calero[9], Mark Rubin[9], Mélanie Tichet[1,3,10], Krisztian Homicsko[1,3,10,11,12], Douglas Hanahan[1,3,10,12]*

Many human cancers manifest the capability to circumvent attack by the adaptive immune system. In this work, we identified a component of immune evasion that involves frequent up-regulation of fragile X mental retardation protein (FMRP) in solid tumors. FMRP represses immune attack, as revealed by cancer cells engineered to lack its expression. FMRP-deficient tumors were infiltrated by activated T cells that impaired tumor growth and enhanced survival in mice. Mechanistically, FMRP's immunosuppression was multifactorial, involving repression of the chemoattractant C-C motif chemokine ligand 7 (CCL7) concomitant with up-regulation of three immunomodulators—interleukin-33 (IL-33), tumor-secreted protein S (PROS1), and extracellular vesicles. Gene signatures associate FMRP's cancer network with poor prognosis and response to therapy in cancer patients. Collectively, FMRP is implicated as a regulator that orchestrates a multifaceted barrier to antitumor immune responses.

Cancers frequently develop the ability to evade destruction by the immune system. Therapeutic strategies aimed at disrupting barriers to tumor immunity are producing noteworthy results and, in some cases, cures for patients with certain cancers. Most patients, however, do not respond to immunotherapies, or the benefit is only transient (*1*–*3*), which highlights the importance of understanding the cellular and molecular mechanisms underlying immune responses to cancer.

We describe the discovery and characterization of an immune barrier that is diversely erected in the tumor microenvironments (TMEs) of human and experimental tumors as a consequence of up-regulating the fragile X mental retardation protein (FMRP) (encoded by the *FMR1* gene). FMRP is an RNA binding translational regulatory protein, identified as a regulator of synaptic plasticity and architecture, which is an essential component of neuronal function (*4*).

FMRP is expressed at high levels in neuronal synapses and at variably lower levels in organs throughout the body (https://www.proteinatlas.org/ENSG00000102081-FMR1/tissue#top). Its genetic or epigenetic inactivation is the primary cause of the neurodevelopmental disorder fragile X syndrome (FXS), which has instigated extensive investigations into the molecular biology of this protein. Historically, FMRP had been established as a translational repressor (*5*–*7*) and, more recently, had been recognized as a regulator of gene expression, in particular by modulating abundance (through transcription and stability) of hundreds of cellular mRNAs that play roles in neuronal function (*7*, *8*). Although mice and humans lacking FMRP are viable and ostensibly normal beyond prominent defects in the central nervous system (*9*–*11*), the role(s) and functional importance of lower-level FMRP expression in other organs are not well defined. Notably, there have been descriptive clues (*12*–*14*) but no substantive documentation that FMRP might play an immunoregulatory role for T cells.

Several studies have previously described the involvement of FMRP in tumor progression, in particular stimulating invasion and metastasis in certain tumor types (*15*–*17*), which led us to further assess its functional roles through gene knockout (KO) and gene knockdown (KD) in cultured cancer cells and tumors growing in mice. Notably, our research design included the use of FMRP-expressing mouse cancer cell lines that could be established as tumors in immunocompetent mice to evaluate the effects both of FMRP expression and of its attenuation or absence on the heterotypic cell types populating the TME, including potential interfaces with the adaptive immune system.

[1]Swiss Institute for Experimental Cancer Research (ISREC), School of Life Sciences, Swiss Federal Institute of Technology Lausanne (EPFL), 1015 Lausanne, Switzerland. [2]Opna Bio SA, Biopole, 1066 Epalinges, Lausanne, Switzerland. [3]Agora Cancer Research Center, 1011 Lausanne, Switzerland. [4]Swiss Institute of Bioinformatics (SIB), 1015 Lausanne, Switzerland. [5]Institute of Pathology, University of Bern, 3008 Bern, Switzerland. [6]Institute of Bioengineering, School of Engineering, Swiss Federal Institute of Technology Lausanne (EPFL), 1015 Lausanne, Switzerland. [7]Department of Computational Biology, University of Lausanne, 1015 Lausanne, Switzerland. [8]Beijing Pediatric Research Institute, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing 100045, China. [9]Department for BioMedical Research, University of Bern, 3008 Bern, Switzerland. [10]Lausanne Branch, Ludwig Institute for Cancer Research, 1011 Lausanne, Switzerland. [11]Department of Oncology, University Hospital of Lausanne (CHUV), 1011 Lausanne, Switzerland. [12]Swiss Cancer Center Leman (SCCL), 1011 Lausanne, Switzerland.
***Corresponding author. Email: douglas.hanahan@epfl.ch**
†These authors contributed equally to this work.

Results

To further substantiate the previously reported up-regulation of FMRP in tumors (*15*–*19*), we assessed FMRP expression in human tumor tissue microarrays (TMAs) and in cognate de novo mouse tumors from the pancreas, colon, breast, and melanomas compared with normal tissue. Consistent with the aforementioned studies, FMRP is up-regulated in major subsets of human pancreatic ductal adenocarcinoma (PDAC), colon carcinoma, and triple-negative breast cancer (TNBC) (Fig. 1, A, C, and E) as well as in genetically engineered mouse models of PDAC, colon cancer, breast cancer, and melanoma (Fig. 1, B, D, and F, and fig. S1, A and B). By contrast, it is lowly or not detectably expressed in the cognate normal human and mouse tissues (left panels of Fig. 1, A to G; fig. S1, A and B; and fig. S2). FMRP is also up-regulated in a majority of human hepatocellular carcinomas (HCCs) and all analyzed primary and brain-metastatic prostate cancers, in contrast to normal or benign tissue (Fig. 1, G and H). An independent investigation of cancer patients by the Human Protein Atlas (*20*) reported that FMRP is variably up-regulated in most forms of human solid tumors (fig. S1C). In particular, we found that FMRP was highly up-regulated in most but not all tumor samples represented in the analyzed TMAs (boxplots in Fig. 1, A, C, E, G, and H). To substantiate the magnified fields shown in Fig. 1, the entirety of representative 0.6-mm cores from the analyzed human TMAs are shown in fig. S2, illustrating a normal tissue core, three FMRP-positive tumor cores, and one FMRP-negative tumor core. Analysis of the largest patient cohort profiled by the Cancer Genome Atlas (TCGA) (*21*) revealed no evidence that the observed up-regulation of FMRP in human tumors commonly involves mutation or amplification of the *FMR1* gene that encodes it (fig. S1, D and E). This might suggest that FMRP's up-regulation is directly or indirectly mediated by the neoplastic state.

We have previously shown that glutamate-stimulated *N*-methyl-D-aspartate receptor (NMDAR) signaling was one such inducer of FMRP expression (*15*). However, the NMDARs are only up-regulated in a subset of tumor types compared with FMRP (fig. S3, A and B), which implicates other regulatory mechanisms in cancers. To explore whether transcriptional regulation of FMRP is involved in its up-regulation in cancer, the ENCODE database (*22*) of chromatin immunoprecipitation sequencing (ChIP-seq) analyses for transcription factors (TFs) binding

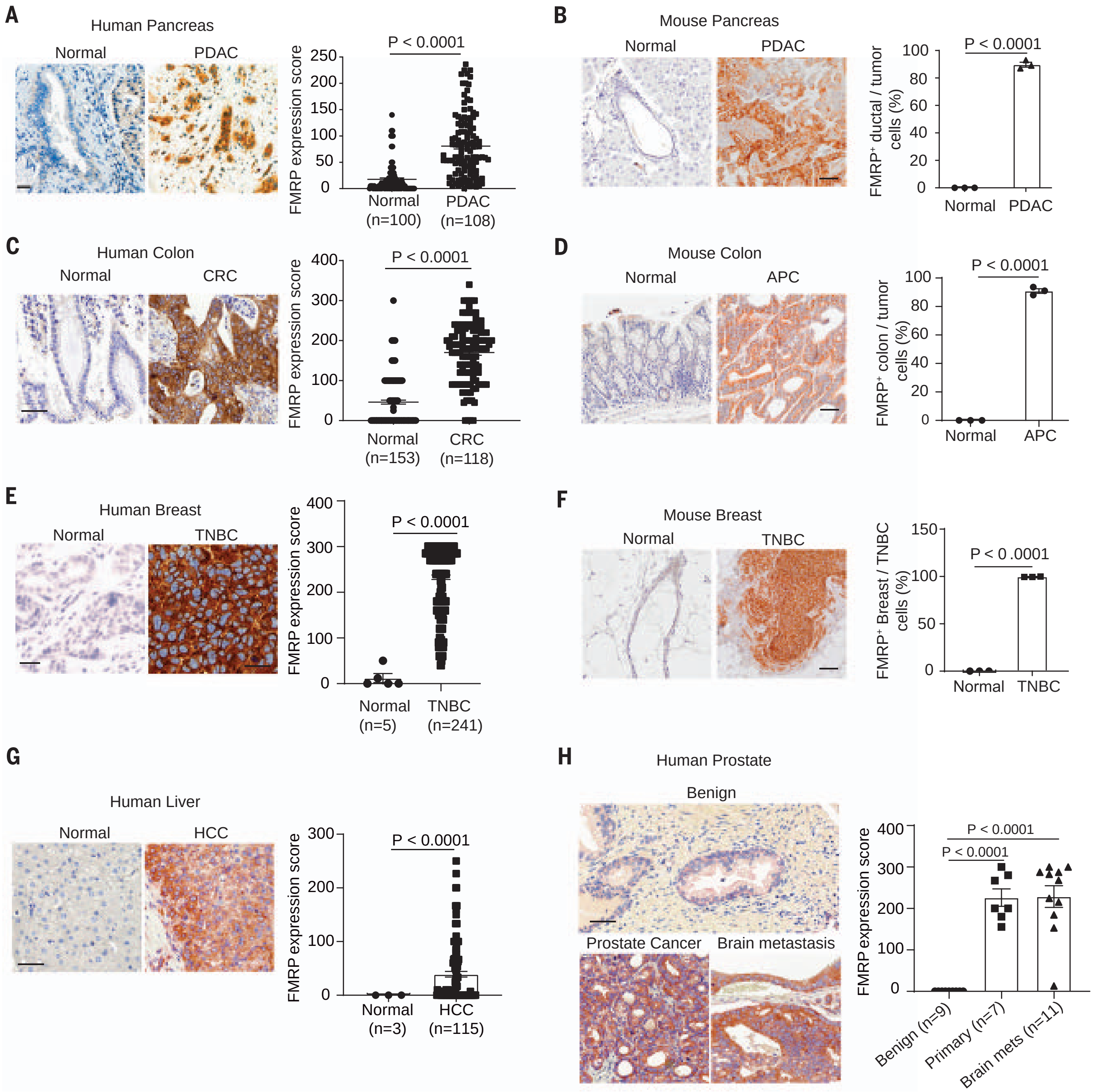


Fig. 1. FMRP is up-regulated in diverse human tumors and cognate de novo mouse tumors. Representative IHC images and quantification of FMRP expression in the indicated human and mouse tumor tissues and normal tissue controls. For the human TMAs, the FMRP expression score = [percentage of FMRP-positive cells in an entire core section (0 to 100) × staining intensity (0 to 3)]. See the Materials and methods for details; scoring was performed independently by two individuals and reconciled. Data are presented as means ± SEMs. (**A**) Human PDAC tumors and tumor-adjacent normal human pancreas in 0.6-mm-diameter core sections of a TMA. N = 108 PDAC tumors and N = 100 normal pancreases were analyzed. Statistics were by Mann-Whitney test. Scale bar, 50 μm. (**B**) Mouse PDAC tumors and normal pancreas from the *P48-cre; LSL-Kras*^{G12D}*; P53*$^{R172H/+\ PDAC}$ mouse model in the FVB/N background (*82*). N = 3 mice per group. For each mouse, five fields from two tissue sections from one independent tumor were analyzed. Statistics by unpaired t test. Scale bar, 50 μm. (**C**) Human colorectal tumors (CRCs) in tumor-adjacent normal colon and in a TMA, analyzed as in (A). N = 118 CRCs and N = 153 normal colons. Statistics by Mann-Whitney test. Scale bar, 50 μm. (**D**) Mouse colon tumors and normal colon from the *ApcΔ/Δ; CDX2 Cre ERT2* (APC) mouse model in the FVB/N background (*83*) For each mouse (N = 5 per group), five fields from two tissue sections from one independent tumor were analyzed. Statistics by unpaired t test. Scale bar, 50 μm. (**E**) Human triple-negative breast cancer (TNBC) in a TNBC TMA (N = 241) and normal breast tissue sections (N = 5, from reduction mammaplasties), analyzed as in (A). Statistics by Mann-Whitney test. Scale bar, 50 μm. (**F**) Mouse breast and TNBC-like breast tumors arising in the C3-Tag mouse model (*84*) in the FVB/N background. For each mouse (N = 5 per group), five fields from two tissue sections from one independent tumor were analyzed. Statistics by unpaired t test. Scale bar, 50 μm. (**G**) Human hepatocellular carcinoma (HCC) and normal liver in a TMA; N = 115 HCC tumor cores and N = 3 normal livers cores; analyzed as in (A). Statistics by one-sample Wilcoxon test. Scale bar, 50 μm. (**H**) Human primary prostate carcinoma, brain-metastatic prostate carcinoma, and normal or benign human prostate tissue in a TMA. N = 7 primary prostate tumor cores, N = 11 brain-metastatic prostate tumor cores, and N = 8 normal pancreas cores; analyzed as in (A). Statistics by one-way analysis of variance (ANOVA) with Kruskal-Wallis test. Scale bar, 50 μm.

to DNA elements was queried for ChIP-seq peaks in the FMR1 promoter region (see Materials and methods). The MYC TF, one of the most widely up-regulated oncogenes across different human cancers (*23*) (fig. S3C), as well as its transcriptional regulatory partner MAX were among the top TFs with significant peaks at the *FMR1* promoter in a variety of cancer cell lines (fig. S3, D and E). We next queried a recently published dataset (*24*) involving a switchable genetic mouse model of PDAC, whereby *Myc* expression can be activated in indolent $Kras^{G12D}$-induced pancreatic intraepithelial neoplasias (PanINs), which results in the progression of PanIN to adenocarcinoma (fig. S3F). Consistently, upon *Myc* activation, *Fmr1* mRNA was significantly up-regulated in PDAC tumors (fig. S3, G and H), encouraging the hypothesis that MYC is a regulator of FMRP expression in cancer. Additionally, to further assess this possibility, we evaluated *Fmr1* mRNA as well as FMRP protein expression levels upon *Myc* KD with small interfering RNA (siRNA) in an FMRP-expressing mouse PDAC cancer cell line, further described below. Concordantly, *Myc* KD resulted in a significant reduction of FMRP expression both at mRNA and protein levels (fig. S3, I to K). Collectively, these data implicate the MYC oncogene in the broad up-regulation of FMRP across the spectrum of solid tumors (fig. S3).

Assessing FMRP functions by its genetic deletion in cancer cells

Aiming to extend upon previous functional studies involving short hairpin RNA (shRNA) KD of *FMR1* in cancer cells (*15*–*19*), we used CRISPR-Cas9 technology to genetically delete the *FMR1* gene in a series of single-cell cloned mouse cancer cell lines, beginning with the PDAC line 4361.12 (*15*) and the colon cancer line CT26 (*25*). For the analyses, we selected two PDAC *Fmr1-KO* clones and one CT26 *Fmr1-KO* clone that were devoid in expression of FMRP protein (fig. S4, A, B, and I). The FMRP-KO had no discernable effect on proliferative phenotypes of the PDAC or CT26 cancer cells in culture compared with their otherwise isogenic wild-type (WT) progenitors (fig. S4, C to E, J, and K). The levels of cell viability (fig. S4D) and cell proliferation (fig. S4E) were similar; the modestly increased rate of apoptosis seen in culture with the KO2 line compared with the WT and KO8 lines (fig. S4F) is not recapitulated in tumor growth assays (fig. S4G) and hence was judged inconsequential. Next, confirming the results of the *Fmr1* siRNA KD analysis performed previously (*15*), the FMRP-KO2 line proved to be severely impaired in a transwell migration assay, a metric of invasiveness (fig. S4H).

To probe the functional effects of the FMRP-KO on tumor phenotypes in vivo, the WT2 and FMRP-KO2 PDAC cell lines were inoculated in parallel into syngeneic and immunodeficient mice. Tail vein injections were performed to seed the lungs followed by survival analysis (Fig. 2A), and subcutaneous (s.c.) inoculations were used to assess tumor growth (Fig. 2D). Unexpectedly, differences in phenotypes were observed in immunodeficient versus immunocompetent mice bearing experimental lung metastases and primary s.c. tumors. The WT2 and FMRP-KO2 PDAC cells had similar metastatic survival (Fig. 2B) and similar primary tumor growth (Fig. 2E) in immunodeficient SCID/beige mice. In marked contrast, both metastatic and primary tumor survival were extended (Fig. 2, C and G) and lung metastatic tumor burden impaired (Fig. 2F and fig. S5, A to D) for the FMRP-KO cells inoculated into syngeneic, immunocompetent FVBN mice. To address the possibility that these differences were KO clone specific, we analyzed a second pair of WT (WT3) and FMRP-KO (KO8) PDAC clones, which revealed similarly impaired tumor growth in immunocompetent mice (Fig. 2, H and I). Analogous phenotypic differences in primary tumor growth were also observed comparing the WT versus FMRP-KO CT26 colon cancer cells (Fig. 2, J to N). Additionally, an *Fmr1-KO* was engineered into the B16-OVA melanoma cell line (fig. S5E) and the 4T1 breast cancer cell line (fig. S5I). Paired isogenic WT and FMRP-KO single-cell clones were evaluated, again revealing impaired tumor growth in immunocompetent mice for the FMRP-KO clones in these cases (fig. S5, F to H and J to L). Collectively, the data demonstrate that tumors derived from FMRP-KO cancer cells grow more slowly in immunocompetent but not in immunodeficient mice, which suggests an involvement of the adaptive immune system.

Evaluating the effects of the adaptive immune system in WT versus FMRP-KO tumors

To assess the possibility that FMRP loss was triggering an adaptive immune response, we analyzed WT and FMRP-KO tumors for the presence of CD8 and CD4 T lymphocytes. The WT PDAC tumors (clones WT2 and WT3) and the WT CT26 colon tumors (WT12) were immune deserts, largely devoid of CD8 T cells (Fig. 3, A and B). By contrast, the isogenic FMRP-KO tumors were highly inflamed by CD8 T cells (Fig. 3, A and B). Fluorescence-activated cell sorting (FACS) analysis of tumors derived from the two sets of WT and FMRP-KO PDAC cell lines confirmed the differential abundance of both CD8 and CD4 T cells (Fig. 3C and fig. S6A). Notably, the CD8 T cells infiltrating the FMRP-KO tumors expressed markers indicative of their activation as cytotoxic T lymphocytes (CTLs) (Fig. 3, D to F, and fig. S6B).

Given that the FMRP-KO tumors were highly inflamed by activated CTLs, we next evaluated their functional contributions to the impaired tumor growth. First, immunocompetent mice bearing PDAC FMRP-KO2 tumors were treated with monoclonal antibodies that deplete CD8 or CD4 T cells (Fig. 3G), which markedly reduced the respective abundance of CD8 and CD4 T cells in the blood (fig. S6, C and D) and of CD8 T cells in the spleen and tumor (fig. S6, E and F). FMRP-KO tumor growth was markedly increased in the mice treated with the depleting anti-CD8 or anti-CD4 antibodies, and more so in combination (Fig. 3H), whereas their survival was commensurately reduced (Fig. 3I), which functionally implicates CD8 and CD4 T cells in the FMRP-KO tumor phenotype. In marked contrast, the T cell–depleting antibodies had no effect on survival or growth of WT2 PDAC tumors (fig. S6G).

Additional evidence for the functional importance of CD8 T cells came from blocking the programmed cell death protein–1 (PD-1) immune checkpoint, based on the observation that WT and FMRP-KO cells and tumors similarly expressed the checkpoint ligand programmed cell death ligand–1 (PD-L1) (fig. S6, H and I). Consistent with the dearth of infiltrating CD8 T cells in mice bearing WT2 tumors, treatment with anti–PD-1 had no effect (Fig. 3J), phenocopying the lack of response in human PDAC (*26*, *27*). By contrast, tumor growth was further impaired when mice bearing FMRP-KO2 tumors were treated with anti–PD-1 antibody, concomitant with inoculation of the cancer cells to seed tumors (Fig. 3K), or after macroscopic solid tumors had formed (Fig. 3L).

To assess possible effects of FMRP-KO cancer cells on the clonal expansion of T cells in tumors, DNA was extracted from WT2 and FMRP-KO2 PDAC tumors and subjected to sequencing analysis for the T cell receptor (TCR) Vß variable region, much as recently described (*28*). The analysis revealed both clonal expansion and markedly increased diversity of TCRs in the FMRP-KO tumors (fig. S7A), consistent with a productive antitumor immune response. Additionally, we assessed the relative immunogenicity of a neoantigen—gp70—expressed in the CT26 colon cancer cell line (*29*). Splenocytes collected from mice bearing tumors derived by s.c. inoculation of WT and FMRP-KO CT26 cell clones were stimulated in culture with gp70 protein, followed by FACS analysis with a gp70-specific tetramer (*30*, *31*). The FMRP-KO tumors elicited increased numbers of tetramer-positive antigen-specific CD8 T cells (fig. S7, B and C), further documenting the immunostimulatory effects of the TME created by cancer cells lacking FMRP.

In addition to its expression in neurons and many cancer cells, FMRP is also expressed in a variety of normal nonneuronal cell types in many tissues (https://www.proteinatlas.org/ENSG00000102081-FMR1/tissue), albeit at generally lower levels than in neurons. Therefore, we assessed possible FMRP expression originating from the stromal compartment of inflamed

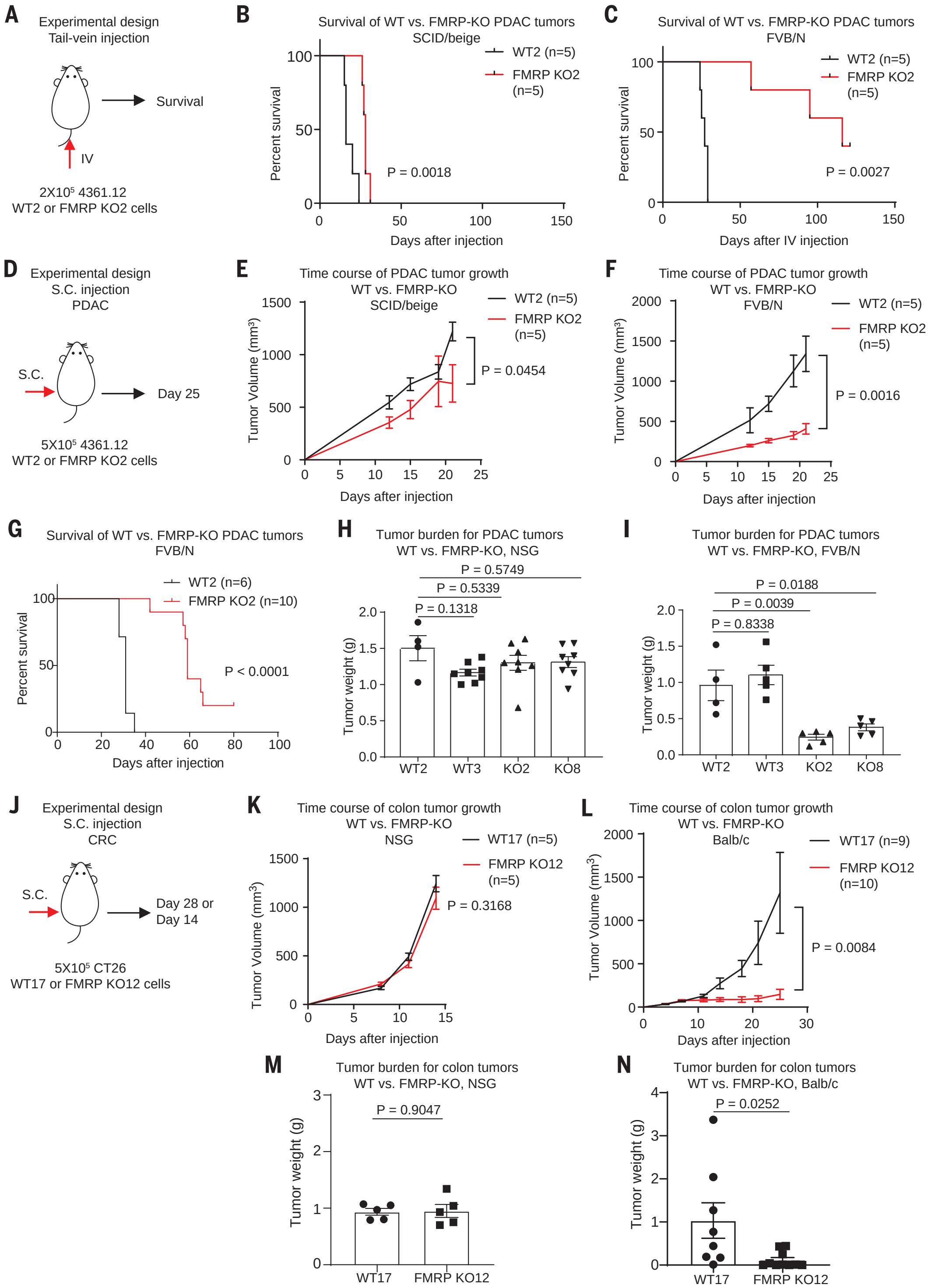


Fig. 2. Genetic deletion of FMRP in cancer cells impairs tumor growth and improves survival in immunocompetent but not immunodeficient mice. (**A**) Schematic of the lung metastasis assay for PDAC WT and FMRP-KO cancer cells (clones WT2 and KO2 from PDAC cell line 4361.12). In total, 2×10^5 cells in 200 μl of PBS were injected into the tail vein of immunodeficient SCID/beige or immunocompetent FVB/N mice. Mice were monitored twice per week for long-term survival. IV, intravenous. (**B** and **C**) Overall survival (OS) of immunodeficient SCID/beige (B) or immunocompetent FVB/N (C) mice bearing lung metastases seeded with PDAC WT or FMRP KO cells. *N* = 5 mice per group; one independent experiment. Statistics by Kaplan-Meier test. (**D**) Schematic of the s.c. primary tumor growth assay after inoculation of PDAC FMRP-WT and FMRP-KO cancer cells (clones WT2 and KO2 from line 4361.12). In total, 5×10^5 cells in 100 μl of PBS were s.c. injected into the right flank of immunodeficient SCID/beige or immunocompetent FVB/N mice. On day 25, all mice were euthanized, and tumors were collected for further analysis. (**E** and **F**) Tumor growth curves of resultant WT and FMRP-KO PDAC tumors in SCID/beige (E) or FVB/N (F) mice. *N* = 5 mice per group. Statistics by unpaired *t* test. Tumor volumes were measured twice per week and calculated using the standard formula, $V = 0.52 \times \text{length} \times \text{width}^2$. Data are representative of two independent experiments and are presented as means ± SEMs. (**G**) OS of immunocompetent FVB/N mice bearing s.c. tumors seeded with PDAC WT2 or FMRP KO2 cells. *N* = 6 mice for the WT group and *N* = 11 for the FMRP KO2 group; one independent experiment. Log-rank test was used for survival analyses. (**H** and **I**) Tumor weights are shown for PDAC WT tumors (formed by two independent clones, WT2 and WT3) and for FMRP-KO tumors (formed by two independent clones, KO2 and KO8) growing in NSG (H) or FVB/N (I) mice at day 25 after s.c. injection. *N* = 4 to 10 mice per group; one independent experiment. Data are presented as means ± SEMs. Statistics by unpaired *t* test. (**J**) Schematic of primary tumor growth assay involving FMRP-WT and FMRP-KO CT26 colon cancer cells. In total, 5×10^5 cells in 100 μl of PBS were s.c. injected into the right flank of immunodeficient NSG or immunocompetent Balb/c mice. On day 14 (for NSG mice) and day 28 (for Balb/c mice), all the mice were euthanized, and tumors were collected for further analysis. (**K** and **L**) Growth curves of resultant WT and FMRP-KO CT26 tumors (from clones WT17 or FMRP KO12) in NSG (K) or Balb/c (L) mice. *N* = 9 mice for the WT17 group and *N* = 10 for the FMRP-KO12 group. Statistics by unpaired *t* test. Tumor volumes were measured twice per week, as above. Data are representative of two independent experiments and are presented as means ± SEMs. (**M** and **N**) Tumor weights for CT26 WT17 versus FMRP-KO12 cells transplanted into NSG (M) or Balb/c (N) mice. Cells were injected s.c., as described in Fig. 2G. *N* = 5 NSG mice for both the WT17 and FMRP-KO12 groups, *N* = 8 Balb/c mice for the WT group, and *N* = 9 Balb/c mice for the FMRP-KO12 group. Statistics by unpaired *t* test. Data are representative of two independent experiments and are presented as means ± SEMs.

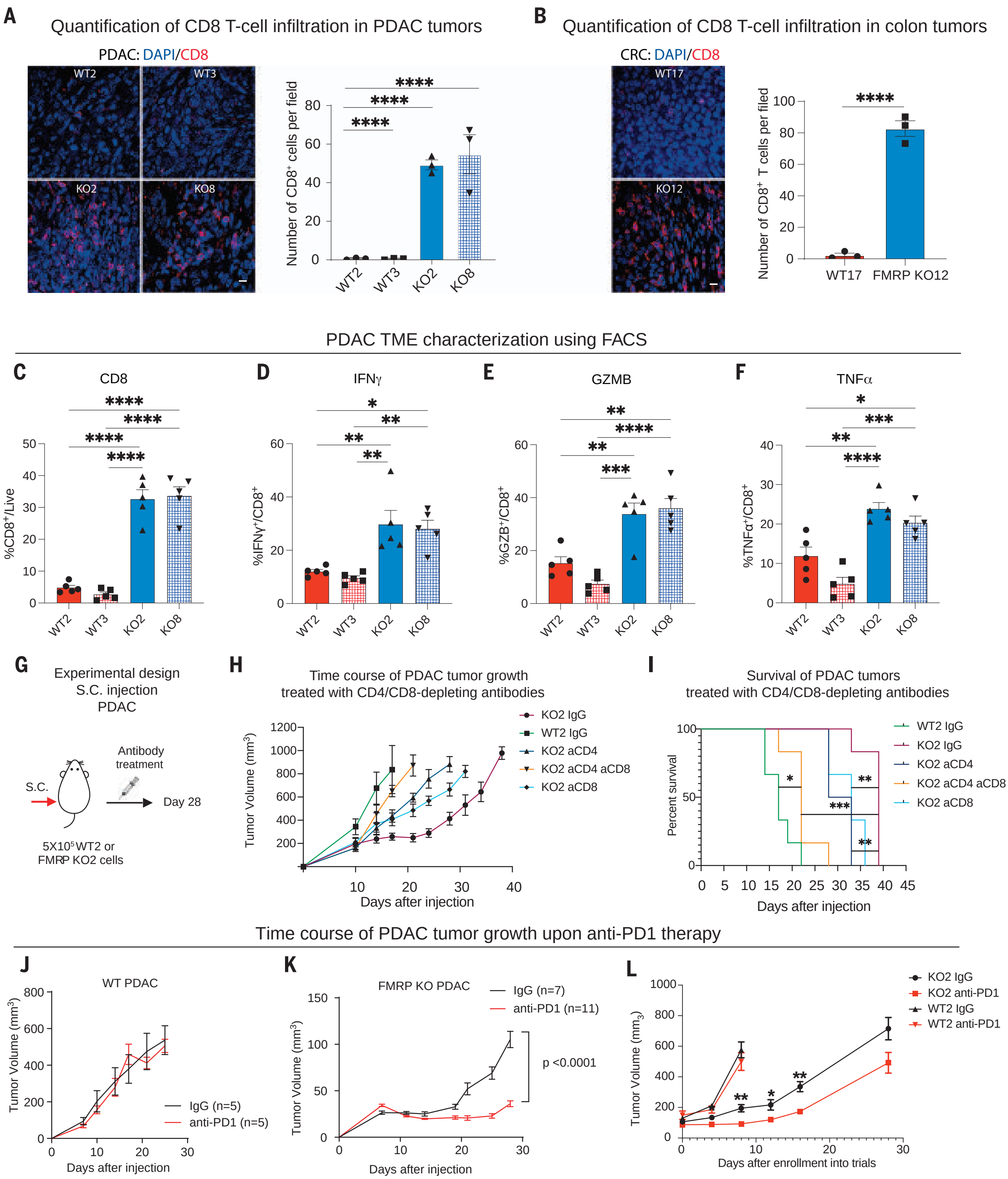


Fig. 3. Tumors lacking FMRP expression in cancer cells elicit inflammation by CD8 and CD4 T cells that are responsible for impaired growth and improved survival. (**A**) Representative IF images and quantification of CD8 T cells in primary s.c. tumors formed by WT and FMRP-KO PDAC cells (clones WT2, WT3, FMRP-KO2, and FMRP-KO8). Scale bar, 10 μm. For each mouse, five fields from two tissue sections from one independent tumor of each mouse were analyzed. Data are presented as means ± SEMs. Statistics by one-way ANOVA with Tukey's multiple comparison analysis. DAPI, 4′,6-diamidino-2-phenylindole. (**B**) Representative IF images and quantification of CD8 T cells in primary s.c. tumors formed by CT26 WT17 and KO12 cells. Scale bar, 10 μm. *N* = 3 mice for each group. For each mouse, five fields from two tissue sections from one independent tumor were analyzed. Data are presented as means ± SEMs. Statistics by unpaired *t* test. (**C** to **F**) FACS analysis of the frequency of $CD11b^{-}CD3^{+}CD8^{+}$ T cells (C) and IFN-γ^{+}– (D), $GZMB^{+}$- (E), and TNF-α^{+}– (F) expressing CD8 T cells infiltrating WT and FMRP KO PDAC tumors from FVB/N mice. *N* = 5 mice for all groups. Data are representative of two independent experiments and are presented as means ± SEMs. Statistics by one-way ANOVA with Tukey's multiple comparison analysis. (**G**) Schematic of the assay to assess the functional importance of CD8 T cells for impaired tumor growth. Mice were injected s.c. with WT2 or FMRP-KO2 PDAC cells, treated with control rat IgG or

rat anti-mouse CD8 or CD4 antibodies, 250 μg per mouse, started on the same day as s.c. injection, through intraperitoneal (IP) injection, twice per week. (**H** and **I**) Tumor growth (H) and survival analysis (I) for FVB/N mice bearing WT2 or FMRP KO2 PDAC tumors, treated with rat IgG or CD4- and/or CD8-depleting antibodies. Mice were euthanized when the tumors reached 1000 mm^3 of volume. N = 6 for all groups. Tumor volumes were measured twice per week. Data are represented as means ± SEMs of one independent experiment. Statistical analysis by Mantel-Cox test. (**J** and **K**) Assessing the response of WT versus FMRP-KO2 tumors to anti–PD-1 immunotherapy. FVB/N mice bearing WT2 (J) or FMRP-KO2 PDAC tumors (K) were treated with control rat IgG antibody or rat anti-mouse PD-1 antibody, and tumor growth was measured twice per week. Mice were euthanized on day 28 after the cancer cell inoculation. N = 5 for both IgG and anti–PD-1 cohorts of WT2 PDAC tumors. N = 7 for the IgG and N = 11 for the anti–PD-1 cohorts of FMRP-KO2 PDAC tumors. Statistics by unpaired t test, Data are presented as means ± SEMs of one independent experiment. (**L**) Intervention trial with the anti–PD-1 antibody in WT2 and FMRP KO2 tumors. Mice were subjected to IgG or anti–PD-1 trials beginning when the tumor volume reached 100 mm^3. Mice were euthanized when the tumors reached 1000 mm^3 of volume or became necrotic. N = 5 for FMRP-KO2 IgG and anti–PD-1 arms and N = 4 for WT2 IgG and anti–PD-1 arms. Statistical analysis by Mann-Whitney test. Data are represented as means ± SEMs of one independent experiment. $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$.

tumors growing in immunocompetent mice derived from inoculation of FMRP-deficient cancer cells. Very low levels of FMRP protein were detected in tumor sections of FMRP-KO PDAC and colon tumors (fig. S7, D and E), whereas no FMRP expression was detected in FMRP-KO breast tumors (fig. S7F), which indicated that the principal source of FMRP detected in WT tumors (Fig. 1 and figs. S1 and S2) is in fact the cancer cells.

The evident differences in immune phenotypes between WT and FMRP-KO tumors raised an ancillary question: Does the resistance versus susceptibility to immune infiltration and attack reflect dominant characteristics of FMRP-WT or FMRP-KO cancer cells? To address this question, we performed mixing experiments whereby WT and FMRP-KO cancer cells were differentially labeled with red and green fluorescent proteins (RFP and GFP) and coinoculated into syngeneic mice at varying ratios. Subsequently, tumor growth, CD8 T cell infiltration, and the localization and abundance of WT and FMRP-KO cells were assessed in tumor tissue sections. Impairment in tumor growth or burden was evident and scaled with the proportion of FMRP-KO cancer cells in the inoculum (fig. S7G), and it was consistently correlated with increased CD8 T cell infiltration (fig. S7H). We focused on the 5:5 mixture with intermediate growth impairment, reasoning that this would best reveal differential and potentially conflicting cellular phenotypes of the two populations of cancer cells. Tumors were collected 7 days after inoculation, and tissue sections were analyzed for green and red fluorescence to visualize the WT and FMRP-KO cancer cells and immunostained with anti-CD8 to identify CD8 T cells. Tumors were highly infiltrated with CD8 T cells in close proximity to both red- and green-labeled cancer cells, revealing that both WT and FMRP-KO cancer cells were associated with the recruited CD8 T cells (fig. S7I). The data indicate that the immunosuppression evident in pure WT tumors cannot be fully conveyed to tumors containing a mixed population of WT and FMRP-KO cancer cells, given the intensive infiltration of CD8 T cells and the similar abundances and interspersion of WT and KO cancer cells in such tumors, which were markedly smaller than those in WT control tumors. Thus, we infer that both WT and FMRP-KO cancer cells can be killed by CTLs, a result further substantiated below in a cell culture bioassay. We also assessed mutational burden and neoantigen load of WT versus FMRP-KO PDAC cells after 30, 60, and 150 days in cell culture by exome sequencing and bioinformatic analysis (Materials and methods), which revealed no significant differences in mutational load or neoantigen abundance (fig. S6, J and K) and hence in inferred immunogenicity.

Finally, to substantiate the FMRP-KO cancer cell phenotype described in the various s.c. transplant models, we evaluated the effects of FMRP in the orthotopic context by implanting FMRP-KO and WT PDAC cells directly into the pancreas of immunocompetent mice. Congruent with the s.c. model, FMRP-KO tumors were significantly smaller and highly infiltrated with CD8 T cells (fig. S6, L to N).

Profiling the cellular constitutions of FMRP-WT versus FMRP-KO tumors through single-cell RNA sequencing

Having established in functional genetic assays that inflammation by CD8 and CD4 T cells of tumors lacking FMRP expression in cancer cells was central to the impaired tumor growth, we sought to investigate the cellular and molecular mechanisms underlying the immunosuppression evident in WT tumors and the adaptive immunity that is triggered in FMRP-KO tumors. As a first step, WT and FMRP-KO PDAC tumors were subjected to single-cell RNA sequencing (scRNA-seq), and clustering analysis was performed to identify the various cell populations in the TMEs. Using published cell type–specific gene signatures (*32*), clusters with similar expression patterns were annotated (fig. S8 and Materials and methods), which revealed marked differences in the cellular constitution of WT versus FMRP-KO tumors (Fig. 4, A and B). The abundance of cancer cells was reduced in FMRP-KO PDAC tumors (Fig. 4C), which, given the unimpaired ability of FMRP-KO PDAC cancer cells to proliferate in culture (fig. S4, C to E) and to grow as tumors in immunodeficient mice (Fig. 2, B and E), implicated the cytotoxicity of T lymphocytes. Congruently, the abundance of infiltrating lymphocytes was markedly increased (Fig. 4D), substantiating the aforementioned analysis that revealed the infiltration of CD8 and CD4 T cells (Fig. 3A) as well as their functional involvement in the impaired growth of FMRP-KO tumors (Fig. 3, H, I, K, and L).

Cluster analysis (Fig. 4, A and B, and fig. S8) revealed less-pronounced variations in other annotated cell types populating the distinctive TMEs (fig. S9A), which led us to focus on further characterization of lymphocytes and to identify subtypes that are differentially abundant in WT versus FMRP-KO tumors. To this end, a computational algorithm for characterizing T cell states and clusters based on reference T cell atlases (ProjecTILs) (*33*) was applied to the lymphocyte cluster in the scRNA-seq data. In addition to the markedly different overall abundance of lymphocytes (Fig. 3A and Fig. 4D), there were appreciable qualitative differences in the fractional representation of lymphocyte subtypes (Fig. 4, E and F, and fig. S9B), most notably the increased fractional (and overall) abundances of effector CD8 T cell subtypes and the reduction in CD4 regulatory T cells (T_{regs}) in FMRP-KO tumors. Motivated by these differences in abundance and in subtype representation, we sought to investigate cell-to-cell signaling from WT and FMRP-KO cancer cells to lymphocytes that might explain these immunological phenotypes.

Charting cell-cell communications in tumors with and without FMRP expression in cancer cells

Using CellChat, an algorithm for mapping cell-cell communications in scRNA-seq datasets (*34*), we assessed differentially increased ligand-receptor pairs in tumors, wherein cancer cells were the "sender cells" expressing genes encoding secreted ligands and the noncancer cells in the TME were the receptor-expressing "recipient cells." Notably, despite the lower abundance of cancer cells in FMRP-KO tumors (Fig. 4C), FMRP-KO cancer cells were implicated by this analysis to have a higher cumulative number of cell-cell communications with other cell types in the TME compared with WT cancer cells (fig. S9C). Given the aforementioned data that indicate profound differences in the

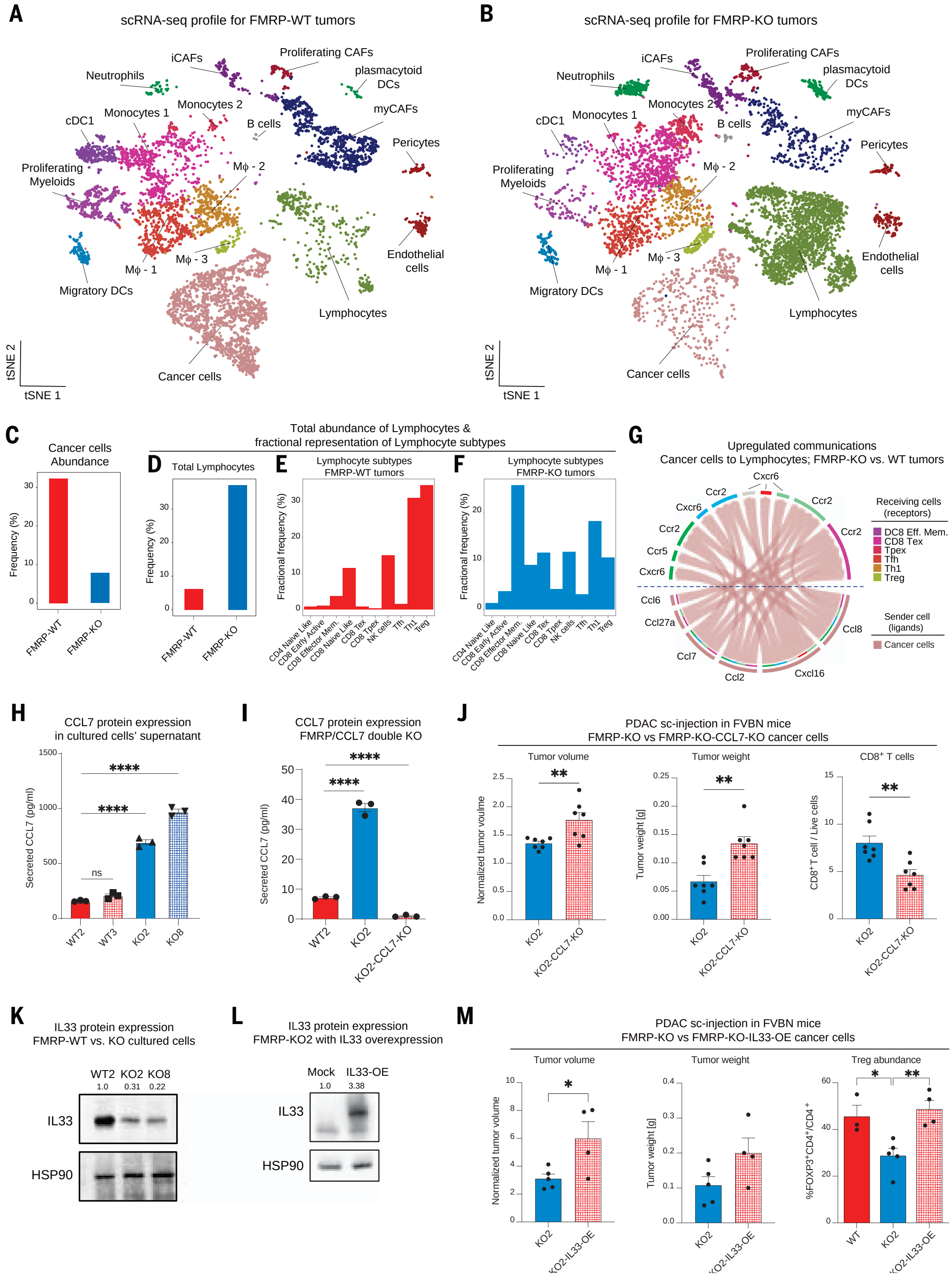


Fig. 4. The absence of upregulated FMRP in cancer cells reprograms the immune TME, as revealed by scRNA-seq. (**A** and **B**) tSNE plots from scRNA-seq analysis for WT tumors (A) and FMRP-KO PDAC tumors (B). The combination of two biological replicates for each population is shown; 18 cell clusters were identified. One independent tumor was analyzed for each cell line (i.e., WT2, WT3, KO2, KO8). CAF, cancer-associated fibroblast; DC, dendritic cell. (**C**) Comparative abundance of the cancer cells in WT (red column) versus FMRP-KO (blue column) PDAC tumors, extracted from the scRNA-seq data. (**D**) Comparative abundance of the lymphocytes in WT (red) versus FMRP-KO (blue) PDAC tumors from scRNA-seq data. (**E** and **F**) Fractional representation of lymphocyte subtypes in WT (E) (red) versus FMRP-KO (F) (blue) PDAC tumors, as revealed by ProjecTILs, an algorithm leveraging reference datasets for T cell subtypes (fig. S8 and Materials and methods). Tfh, T follicular helper cell; Th1, T helper 1 cell. (**G**) CellChat analysis of scRNA-seq data implicating signaling circuits that are up-regulated in FMRP-KO tumors, in particular involving cancer cells as ligand-expressing sender cells and lymphocytes as receptor-expressing receiving cells (see table S1 for specifics). (**H**) ELISA analysis of the levels of CCL7 protein secreted from WT and FMRP-KO PDAC cancer cells in culture. In brief, one million cells were seeded in 10-mm culture plates with complete medium. After 48 hours, culture medium was replaced with 6 ml of serum-free medium. Medium was collected 48 hours later for cytokine profiling and ELISA analysis. *N* = 3 samples for each group. Data are representative of two independent experiments and are presented as means ± SEMs. (**I**) Absence of CCL7 protein expression in FMRP-KO cancer cells engineered to carry a genetic deletion of *Ccl7*. In brief, 2 × 10^5 cells were seeded in 2 ml of complete medium in one well of a 6-well plate. After 48 hours, medium was collected for ELISA analysis of CCL7 protein level. *N* = 3 samples for each group. Data are representative of three independent experiments and are presented as means ± SEMs. (**J**) Comparison of FMRP-KO versus *Fmr1*/CCL7-DKO s.c. tumors. The panels represent tumor volume (left), tumor weight (middle), and FACS analysis of CD8 T cells (right). Mice were euthanized on day 14 after cell implantation. Tumor volumes were normalized between days 7 and 14. CD8 T cells are shown as the proportion of live cells. *N* = 7 mice for both groups; one independent experiment. Statistical analysis was performed with Mann-Whitney test. Data are presented as means ± SEMs. (**K**) Protein blot of IL-33 expression in WT versus FMRP-KO PDAC cell lines. Data are representative of three independent experiments. (**L**) Protein blot demonstrating genetically rescued levels of IL-33 expression in an FMRP-KO cell line. Data are representative of three independent experiments. (**M**) Functional effects of IL-33 overexpression in FMRP-KO tumors. The graphs show means ± SEMs of tumor volume (left), tumor weight (middle), and FACS analysis of T_{regs} (right). Mice were euthanized on day 14 after s.c. injection. Tumor volumes were normalized between days 7 and 14. T_{regs} were defined as $FOXP3^+CD4^+$ cells and shown as a percentage of $CD4^+$ T cells. *N* = 5 for FMRP-KO2 group, *N* = 4 for KO2–IL-33–OE group, and *N* = 3 for WT2 group; one independent experiment. Statistical analysis by Mann-Whitney or one-way ANOVA tests. ns, not significant; *P < 0.05; **P < 0.01; ****P < 0.0001.

lymphocyte abundance and subtype constitution, we sought to chart up-regulated communications from FMRP-KO cancer cells to lymphocytes. Most prominently, a group of chemokines and their receptors were identified that were more frequently expressed in FMRP-KO cancer cells and lymphocytes in the scRNA-seq dataset, respectively (Fig. 4G and table S1). Among the five up-regulated ligands in the scRNA-seq dataset (fig. S9D), only C-C motif chemokine ligand 7 (*Ccl7*) demonstrated significant up-regulation in cultured FMRP-KO cancer cells as both an mRNA (fig. S9E) and a secreted protein (Fig. 4H); congruently, an siRNA KD of *Fmr1* in WT cancer cells also elevated CCL7 secretion (fig. S9F). Moreover, *Ccl7* scored significantly in a previously published FMRP cross-linking immunoprecipitation sequencing (CLIP-seq) analysis (*35*) (fig. S9G) that is indicative of direct binding of mRNAs to FMRP (albeit in neurons), consistent with the implication that FMRP represses *Ccl7* mRNA translation (and mRNA stability) in WT tumors—a repression that is released in cancer cells lacking FMRP expression. Notably, CCL7 is a proinflammatory cytokine capable of contributing to the recruitment of lymphocytes and thereby triggering antitumor immune responses (*36*). To assess the possible role of up-regulated CCL7 in the immune phenotype of FMRP-KO cancer cells, we used CRISPR-Cas9 technology to produce a double knockout (DKO) of the *Fmr1* and *Ccl7* genes in PDAC cancer cells such that CCL7 protein is not expressed (Fig. 4I). Then, we audited these DKO cancer cells upon transplantation compared with FMRP-KO cells. In the absence of cancer cell–derived CCL7, tumor volume and weight were increased compared with those of FMRP-KO tumors, and the abundance of CD8 T cells was reduced (Fig. 4J). We also observed a modest increase in the abundance of T_{regs} but no change in the expression of markers that indicate the immunosuppressive versus proinflammatory phenotypes of tumor-associated macrophages (TAMs) (fig. S9H), which suggests that CCL7 is modulating the inflammation by T cells but not the programming of TAMs. As such, the FMRP-KO phenotype is evidently not singularly modulated by CCL7 because it is not affecting TAMs. Therefore, additional factors expressed by KO cancer cells—including the other four up-regulated chemokines (fig. S9D)—may also be involved, warranting future investigation.

As a complement to the CellChat analysis of the scRNA-seq, bulk RNA-seq was performed for cultured WT and FMRP-KO cancer cells. In this analysis, we focused on the converse question: Namely, are well-recognized immunomodulatory genes comparatively up-regulated in the immunosuppressive FMRP-WT cells? To address this question, the transcriptome datasets for cultured cancer cells were scored for differentially expressed cytokines (table S2), and interleukin 33 (*Il-33*), a chemokine that is known to stimulate the production and activity of T_{regs} (*37*), was identified as the most prominent. *IL-33* mRNA is down-regulated in FMRP-KO cancer cells (fig. S9I), and IL-33 protein level is decreased both in KO and siRNA KD cancer cells (Fig. 4K and fig. S9J). Additionally, CLIP-seq analysis indicated that *Il-33* mRNA binds FMRP (fig. S9K) and is therefore implicated as a direct target of FMRP regulation. To follow up on these observations, an FMRP-KO cancer cell line was engineered that overexpressed IL-33 protein (FMRP-KO–IL-33–OE) (Fig. 4L). After transplantation into syngeneic immunocompetent mice, tumor growth was examined and characterized. Compared with FMRP-KO tumors, the FMRP-KO–IL-33–OE tumors had increased volume and weight and increased abundance of T_{regs}, comparable to WT tumors (Fig. 4M), revealing yet another FMRP-regulated factor involved in modulating antitumor immunity.

Thus, we have identified and functionally validated two direct targets of FMRP expression in cancer cells that are involved through their differential expression in the distinctive effects of WT versus FMRP-KO cancer cells on the immune TME. CCL7 is up-regulated in KO tumors, leading to the recruitment of CD8 T cells, whereas the concomitant reduction in IL-33 expression compared with that in WT tumors led to fewer T_{regs}, consistent with the marked differences in lymphocyte abundance and subtype representation that distinguishes WT and FMRP-KO tumors (Fig. 4, D to F, and fig. S9B).

Assessing interactions of T cells with WT versus FMRP-KO cancer cells in coculture

Given the results on cancer cell–to-lymphocyte cell-cell communication, we next sought to further assess the implication that WT and FMRP-KO cancer cells have direct and differential interactions with CD8 and CD4 T cells. To this end, a coculture assay was established (Fig. 5A and Materials and methods) involving ex vivo activated and tumor antigen–nonspecific splenic T cells, which were combined with WT versus FMRP-KO cells, or parental cancer cells transfected with control siRNA versus siRNA targeting *Fmr1*. Subsequently, T cell proliferation without the confounding complexity of T cell–mediated killing of the cancer cells was assessed through FACS analysis. Notably, because the anti-CD3 plus anti-CD28 bead-activated T cells are already highly proliferative, this assay only reports upon potential inhibitory effects of cancer cells on T cell proliferation. Two pairs of WT and FMRP-KO PDAC cancer cells were assessed, as were two pairs of parental PDAC cells treated with two independent control siRNAs or siRNAs to *Fmr1* (FMRP-KD), each in coculture with CD8 or CD4 T cells. Compared with T cells cultured alone, the WT and siRNA-control (WT) cancer cells were potently inhibitory for CD8 and CD4 T cell proliferation (Fig. 5, B and C). By contrast, the FMRP-KO and FMRP-KD PDAC cells only modestly inhibited proliferation of CD8 (Fig. 5B) and CD4 (Fig. 5C) T cells. Additionally, we compared PDAC WT2 cells and KO2 cells with the double KO of FMRP and CCL7, which revealed that the FMRP-CCL7 DKO cells were similarly inhibitory toward CD8 T cell proliferation as WT cancer cells (Fig. 5D), which further established CCL7 as a key FMRP-regulated factor involved in counteracting the direct immunosuppressive effect of WT cancer cells on CD8 T cells.

To extend on these results, we also compared the effects on proliferation of activated antigen-nonspecific T cells with matched pairs of WT versus FMRP-KO B16-OVA melanoma cells and CT26 colon cancer cells. FMRP-KO B16-OVA cells similarly evidenced reduced inhibitory activity for both CD8 and CD4 T cell proliferation compared with WT B16-OVA cancer cells (Fig. 5, E and F). Concordantly, both FMRP-KO and *siRNA Fmr1* (FMRP-KD) parental CT26 cells showed reduced inhibition compared with WT or *siRNA Ctrl* CT26 cells (Fig. 5, G and H).

Next, we assessed antigen-specific killing by CD8 T cells in a classical CTL assay, involving OVA-specific OT1 CD8 T cells (isolated from OT1 transgenic mice) cocultured with WT or FMRP-KO B16-OVA cancer cells, both expressing similar levels of OVA antigen (Fig. 5I and fig. S9L). A T cell titration assay was performed, which revealed the sensitivity of both WT and FMRP-KO cancer cells to T cell killing, consistent with the aforementioned mixing experiment (Fig. 3, G, H, and I). The FMRP-KO cells were more susceptible in this assay to T cell killing at all ratios (Fig. 5J). The differential sensitivity to T cell killing was confirmed in a second FMRP-KO B16-OVA cell line as well as with matched pairs of siRNA control and siRNA FMRP-KD transfected B16-OVA cancer cells, as shown comparatively at the 1:1 ratio (Fig. 5K).

Thus, FMRP-KO (and KD) cancer cells are (i) markedly less inhibitory toward T cell proliferation, in part resulting from expression of CCL7, and (ii) somewhat more susceptible to CD8 T cell killing. The results clearly establish a direct mechanistic link between the FMRP-deficient cancer cells and the adaptive immune response catalyzed in FMRP-KO tumors.

Distinctive phenotypes of macrophages populating WT versus KO tumors

We next considered another possible parameter that might distinguish the immunosuppressed WT versus the T cell–inflamed FMRP-KO tumors, namely TAMs and myeloid cells, collectively defined by expression of the cell surface marker CD11b. These innate immune cells are well

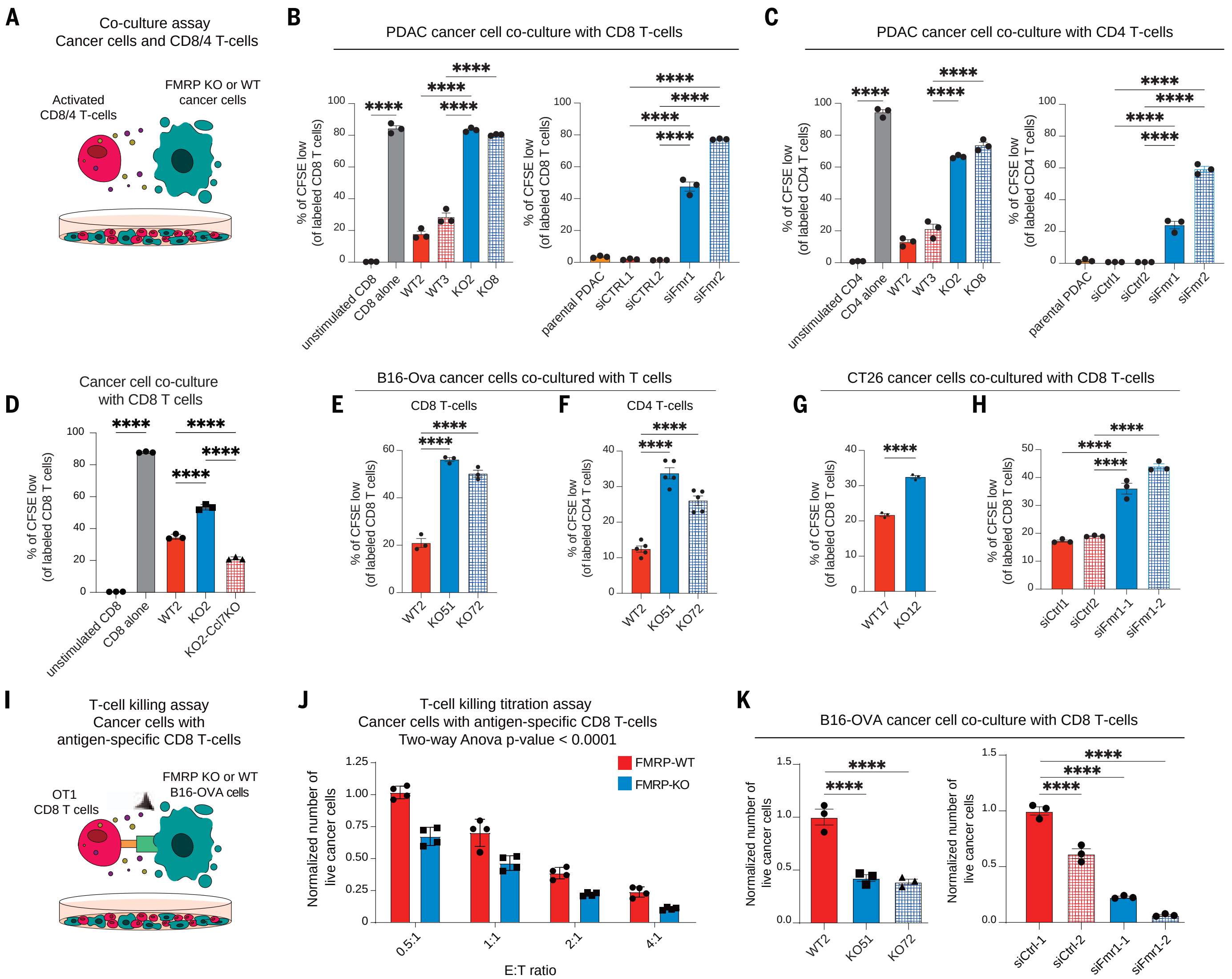


Fig. 5. FMRP-KO cells directly stimulate T cell proliferation and are more sensitive to killing by antigen-specific CD8 T cells. (**A**) Schematic representation of the ex vivo T cell coculture assay. Splenic CD4 or CD8 T cells were prestained with carboxyfluorescein diacetate succinimidyl ester (CFSE), activated with the CD3 or CD28 Dynabeads, and cocultured with FMRP-KO or WT PDAC cells. After 72 hours of coculture, cells were harvested and processed for flow cytometry analysis. (**B** and **C**) FACS analysis of CD8 (B) and CD4 (C) T cell proliferation in cocultures with WT or FMRP-KO PDAC cancer cells (left), or WT2 cells transfected with *siCtrl* or *siFmr1* RNAs (right). The graphs represent the fraction of proliferating (CFSE^{-low}) cells as a fraction of all labeled CD8 or CD4 T cells. *N* = 3 samples for each group. Data are representative of three independent experiments and are presented as means ± SEMs. Statistics by one-way ANOVA with Tukey's multiple comparison analysis. (**D**) FACS analysis of CD8 T cell proliferation in cocultures with WT, FMRP-KO2 PDAC, and KO2-CCL7KO (DKO) PDAC cancer cells. The graph represents the fraction of proliferating (CFSE^{-low}) cells as a fraction of all labeled CD8 T cells. *N* = 3 samples for each group. Data are representative of three independent experiments and are shown as means ± SEMs. Statistics by one-way ANOVA with Tukey's multiple comparison analysis. (**E** and **F**) FACS analysis of CD8 (E) and CD4 (F) T cell proliferation in cocultures with WT or FMRP-KO B16-OVA cancer cells. The graphs represent the fraction of proliferating (CFSE^{-low}) cells over all labeled CD8 or CD4 T cells. *N* = 3 samples for each group. Data are means ± SEMs and are representative of two independent experiments. Statistics by one-way ANOVA with Tukey's multiple comparison analysis. (**G** and **H**) FACS analysis of CD8 T cell proliferation in cocultures with WT or FMRP-KO CT26 cancer cells (left), or parental CT26 cells transfected with *siCtrl* or *siFmr1* RNAs (right). The graphs represent the fraction of proliferating (CFSE^{-low}) cells as a fraction of all labeled CD8 or CD4 T cells. *N* = 3 samples for each group. Data are representative of two independent experiments and represent means ± SEMs. Statistics by one-way ANOVA with Tukey's multiple comparison analysis. (**I**) Schematic representation of an in vitro T cell killing assay. B16-OVA WT and FMRP-KO cells were cocultured in different ratios with CD8^{+} T cells isolated from OVA-specific T cell receptor transgenic (OT1) mice that had been preactivated by OVA peptide and IL-2 plus IL-7 treatment. The number of viable B16-OVA cancer cells at the end point of the assay was determined and reported. (**J**) Titration assay involving coculture at different effector-to-target ratios (E:T ratios) assessing sensitivity to cytotoxic T cell killing of B16-OVA WT and FMRP-KO cells; after 48 hours, FACS analysis revealed the numbers of live cancer cells, which were normalized to the WT cancer cell number in the 0.5:1 group. *N* = 4 samples for each group. Statistics by two-way ANOVA test. Data are representative of two independent experiments and represent means ± SEMs. (**K**) Further validation of the T cell killing assay using OT1 CD8 T cells and B16-OVA cancer cells at a 1:1 ratio. (Left) Two independent B16-OVA FMRP KO clones (KO51 and KO72) after 48 hours of coculture; FACS analysis revealed the numbers of remaining cocultured cancer cells, which were normalized to the WT cancer cell number. (Right) B16-OVA WT cells transfected with *siCtrl* or *siFmr1* RNAs were analyzed after 24 hours. FACS analysis revealed the numbers of cocultured cancer cells, normalized to the *siCtrl1* group. Statistical analysis by one-way ANOVA. *N* = 3 samples for each group. Data are representative of three independent experiments and represent means ± SEMs. ****P < 0.0001.

known to populate PDAC and other solid tumors, in particular TAMs displaying a so-called M2-like immunosuppressive phenotype (38, 39). To begin, we modified the T cell proliferation assay (Fig. 5A), replacing the cancer cells with myeloid cells isolated by immunomagnetic selection for CD11b-positive cells from tumors established with the two matched pairs of PDAC WT and FMRP-KO cells. The CD11b$^+$ myeloid cells isolated from both WT tumors suppressed CD8 and CD4 T cell proliferation, whereas those from FMRP KO tumors did not discernibly inhibit T cell proliferation, similar to controls without added myeloid cells (Fig. 6, A and B). Additionally, we further assessed the phenotype of tumoral macrophages by FACS from one matched set of WT and FMRP-KO tumors (PDAC WT2 and KO2) for expression of CD86, a marker of M1-like immunostimulatory macrophages, and arginase 1 (ARG1), a marker of M2-like immunosuppressive macrophages. Their converse levels of expression suggested that WT tumors had more abundant M2-like macrophages, whereas FMRP-KO tumors favored M1-like TAMs (Fig. 6C). These results collectively motivated investigation of the possibility that the WT versus FMRP-KO cancer cells were differentially programming TAMs.

A transwell coculture assay was established (Fig. 6D) and used to assess the effects of WT versus FMRP-KO cancer cells on bone marrow–derived macrophages (BMDMs), which were programmed ex vivo to adopt an M1-like phenotype (Materials and methods). WT cancer cells induced the expression of two immunosuppressive proteins—ARG1 and IL-10—prototypically expressed in M2-like macrophages, whereas the FMRP-KO cancer cells slightly induced ARG1 and did not affect IL-10 (Fig. 6E). FACS analysis of the M1-polarized BMDMs from the transwell assay confirmed the more-pronounced induction of ARG1 by WT versus FMRP-KO cancer cells and further revealed that WT cancer cells down-regulated expression of the M1-like marker CD86, whereas the FMRP-KO cells had no effect (Fig. 6F), in agreement with the analysis of these two markers in bona fide TAMs (Fig. 6C). Additionally, mRNA levels of four immunosuppressive genes expressed in M2-like TAMs were preferentially up-regulated in M1-polarized BMDMs by WT cancer cells in the transwell coculture assay (Fig. 6G). The results with ARG1 as a biomarker of the M2-like TAM state induced by WT cancer cells were further corroborated in the analysis of a second pair of WT versus FMRP-KO cancer cells (fig. S10A) and in a comparison of siRNA control versus siRNA FMRP-KD cells (fig. S10B). Collectively, these bioassays reveal that FMRP-expressing WT cancer cells are differentially active in their capability to reprogram M1-polarized macrophages into an M2-like immunosuppressive phenotype, consistent with the distinctive immune deserts that characterized FMRP-WT tumors (Fig. 3, A and B).

Given the evident M2-like TAM programming activity of FMRP-WT cancer cells, we sought to identify differentially expressed factors that might contribute to their induction. Returning to the RNA-seq comparison of WT and FMRP-KO cancer cells in culture (table S2), a mechanistically logical gene—*Pros1*—was found to be preferentially expressed in WT cancer cells. Tumor-secreted protein S (PROS1), a ligand for the MER/TYRO3 receptor that is expressed on macrophages, has been shown to repress expression of immunostimulatory genes in M1-like macrophages, and its genetic ablation in cancer cells enhances antitumor immune responses (40). Our investigation revealed that PROS1 protein and mRNA are more highly expressed in WT versus FMRP-KO cancer cells (Fig. 6H and fig. S10C), and CLIP-seq analysis indicated that *Pros1* mRNA can bind specifically to FMRP, implicating it as a translational (and transcriptional) regulatory target (fig. S10D). To assess its possible functional involvement, we first cocultured M1-polarized BMDMs in the transwell assay with WT versus FMRP-KO PDAC cancer cells in the presence of recombinant mouse PROS1 protein. FMRP-KO cells, which weakly induced immunosuppressive ARG1 compared with WT cancer cells (Fig. 6E and fig. S10A), elevated ARG1 expression in macrophages to a level comparable to that of WT cancer cells in coculture with M1-polarized BMDMs when both were similarly treated with PROS1 protein (Fig. 6I). Concordantly, the addition of PROS1 to the coculture medium with FMRP-KO cancer cells served to inhibit expression of the aforementioned M1-like marker genes, *Cd86* and *Il-1α*, to the same level as with WT cells (Fig. 6J).

Next, we overexpressed PROS1 protein in FMRP-KO cancer cells (fig. S10E) and assessed phenotypic effects in both the transwell assay with M1-polarized BMDMs and tumors. Forced expression of PROS1 protein in FMRP-KO cancer cells resulted in marked up-regulation of ARG1 protein (fig. S10F) and down-regulation of *Cd86* and *Il-1α* mRNA (fig. S10G) in the transwell assay, congruent with the results involving the addition of PROS1 protein (Fig. 6, I and J). Moreover, when FMRP-KO (KO2) and PROS1-overexpressing FMRP KO (*KO2–Pros1-OE*) PDAC cancer cells were transplanted into syngeneic mice, up-regulation of PROS1 protein led to modestly increased tumor growth, tumor weight, and ARG1 expression as well as reduced expression of CD86 in TAMs (fig. S10H). These results suggest another mechanistic component of FMRP's mode of action in immune evasion, whereby FMRP directly regulates PROS1 expression in WT cancer cells, thereby contributing to an immunosuppressive TME through its induction of M2-like TAMs, such that its down-regulation in FMRP-KO cancer cells facilitates the generation of immunostimulatory M1-like macrophages.

Upon curating the scRNA-seq data for other potentially immunosuppressive factors in WT tumors, *Cd63* mRNA was identified as one of the most differentially expressed genes in WT versus FMRP-KO cancer cells (Fig. 6K and table S2); notably, *Cd63* encodes a protein that is displayed on and involved in the secretion of exosomal vesicles (EVs) (41). Investigation of exosome secretion from cultured WT versus FMRP-KO cancer cells revealed that the WT cells secreted more exosomes in both the PDAC and B16-OVA matched pairs (Fig. 6L and fig. S10I), consistent with the higher levels of *CD63* mRNA. Notably, a recent study revealed a direct role for FMRP in the process of exocytosis during inflammation (42), which suggests that the regulation of exosome formation and secretion might be another mechanism by which FMRP mediates its immunosuppressive effects.

Therefore, we reasoned that the secreted WT cancer cell–derived exosomes might be involved in macrophage programming, as has been suggested in previous studies (43, 44). Accordingly, EVs were isolated from WT and FMRP-KO PDAC cancer cells (fig. S10J) and added in equal amounts to the cultured M1-polarized BMDMs. WT exosomes markedly up-regulated mRNAs for four immunosuppressive genes characteristically expressed in M2-like TAMs (Fig. 6M), whereas FMRP-KO exosomes had little or no effect (Fig. 6M), which was also apparent at the protein level for ARG1 (fig. S10K). To further assess the differential effects of WT versus FMRP-KO exosomes, we polarized BMDMs to an M2 phenotype and cultured them in the presence of exosomes. WT exosomes further increased the expression of *Chil3* and *Arg1* mRNA in M2-polarized BMDMs (fig. S10L) and slightly increased the already appreciable ARG1 protein levels (fig. S10M); by contrast, FMRP-KO exosomes had no effect on *Chil3* mRNA, modestly increased *Arg1* mRNA, and nevertheless reduced ARG1 protein levels (fig. S10, L and M). Therefore, the exosomes secreted by FMRP-KO cancer cells were both less abundant and functionally distinct given that—even when applied in equal amounts—the EVs from FMRP-KO cancer cells had markedly reduced capability to induce immunosuppressive gene expression in phenotypically polarized BMDMs.

Both PROS1 and exosomes produced by FMRP-expressing cancer cells are directly regulated by FMRP and implicated in the programming of an immunosuppressive TME through the induction of M2-like TAMs. In the absence of FMRP, the TAMs are evidently biased toward an M1-like immunostimulatory phenotype that is expected to contribute to the above-documented recruitment and activation of CD8 and CD4 T cells.

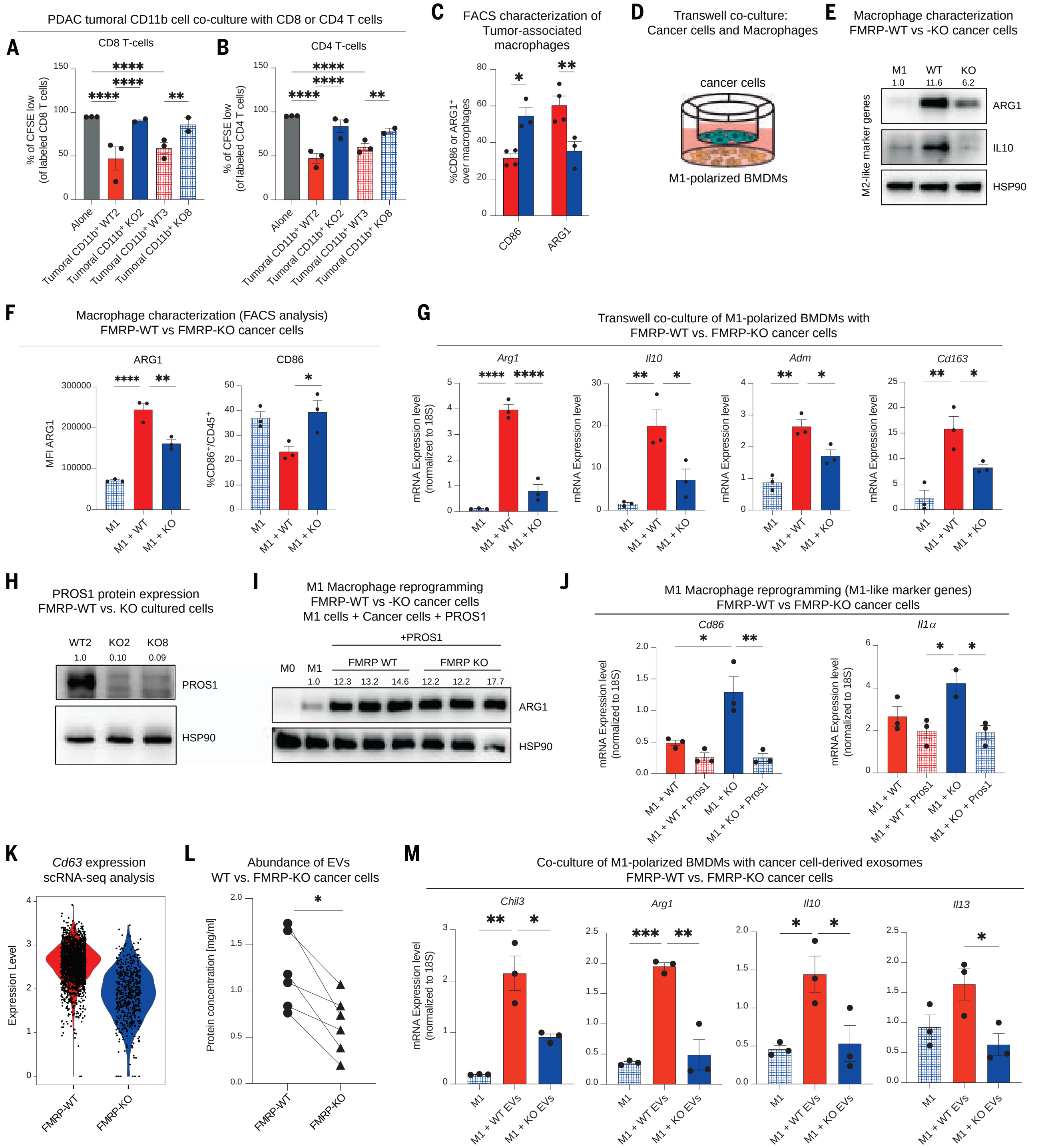


Fig. 6. Immunosuppressive M2-like macrophages are abundant in the TME of WT but not FMRP-KO tumors, programmed by PROS1 and cancer cell–derived exosomes that are both more highly expressed by WT cancer cells. (**A** and **B**) FACS analysis of immunosuppressive effects of tumoral CD11b cells—isolated by FACS from WT (WT2 and WT3) or FMRP-KO (KO2 and KO8) PDAC tumors—on CD8 (A) or CD4 (B) T cell proliferation after 72 hours of coculture. Proliferation is shown as the percentage of CFSE$^{\text{-low}}$ cells compared with the total number of CD8 or CD4 T cells. N = 2 to 3 samples for each group. Statistical analysis by one-way ANOVA. Data are representative of two independent experiments and are presented as means ± SEMs. (**C**) FACS analysis of CD86 (M1 TAM–like marker) or ARG1 (M2 TAM–like marker) expression in macrophages from WT and FMRP-KO PDAC tumors. N = 4 for WT tumors and N = 3 for KO tumors. Data are presented as means ± SEMs. Statistical analysis by Mann-Whitney test; one independent experiment. (**D**) Schematic representation of a transwell coculture assay involving ex vivo programmed M1-polarized BMDMs combined with WT or FMRP-KO cancer cells. (**E**) Protein blot analysis of M2 TAM–like marker (ARG1 and IL-10) expression in M1-polarized BMDMs cocultured with WT or FMRP-KO cancer cells. Data are representative of three independent experiments. (**F**) FACS analysis for ARG1 and CD86 (M1 TAM–like marker) protein expression for M1-polarized BMDMs cocultured with WT or FMRP-KO cancer cells. N = 3 samples for each group. Data are representative of two independent experiments and are shown as

means ± SEMs. (**G**) RT-PCR quantification of M2 TAM–like marker gene (*Chil3*, *Arg1*, *Il-10*, *Il-13*) mRNA expression in M1-polarized BMDMs cocultured with cancer cell–derived exosomes from WT or FMRP-KO cancer cells. *N* = 3 samples for each group. Data are representative of two independent experiments. Error bars represent means ± SEMs. (**H**) Protein blotting analysis of PROS1 expression in WT versus FMRP-KO PDAC cancer cells. Data are representative of three independent experiments. (**I**) Western blotting analysis of ARG1 protein expression in M1-like macrophages cocultured with WT or FMRP-KO cancer cells, with or without addition of recombinant PROS1 protein to the culture media. *N* = 3 samples for each group. Data are representative of two independent experiments and are shown as means ± SEMs. (**J**) RT-PCR quantification of M1 TAM–like marker gene (*Cd86*, *Il-1α*) mRNA expression in M1-polarized BMDMs cocultured with WT or FMRP-KO cancer cells, with or without PROS1 protein in the media. *N* = 3 samples for each group. Data are representative of two independent experiments and are presented as means ± SEMs. (**K**) Violin plot showing the distribution of mRNA expression for the prototypical extracellular vesicle marker, *Cd63*, in WT versus FMRP-KO tumors. See table S2 for the fold change and level of significance. (**L**) Comparative abundance of isolated extracellular vesicles from WT and FMRP-KO PDAC cancer cells in culture. Pooled data of six independent isolation experiments with six paired WT and FMRP KO samples. Statistical analysis by paired *t* test. (**M**) RT-PCR quantification of M2 TAM–like marker gene (*Chil3*, *Arg1*, *Il-10*, *Il-13*) mRNA expression in M1-polarized BMDMs cocultured with cancer cell–derived exosomes from WT or FMRP-KO PDAC cancer cells. Statistical analysis by one-way ANOVA. *N* = 3 samples for each group. Data are representative of two independent experiments and are shown as means ± SEMs. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$.

Contributions of macrophages to tumor immunity in FMRP-KO tumors

Given these implications, we next sought to assess indirectly regulated components of the FMRP-KO immunostimulatory phenotype, in particular the putative contributions of M1-like immunostimulatory TAMs. First, we performed a CellChat analysis of the scRNA-seq datasets to identify potential cell-to-cell communications preferentially up-regulated in FMRP-KO tumors in the form of genes for secreted ligands that were expressed in myeloid cells and macrophages (as senders) versus cognate receptors that were expressed in lymphocytes (as receivers) (table S1). Among the comparatively elevated signaling interactions in FMRP-KO tumors, most belonged to the CCL and CXCL families of cytokines (Fig. 7A), which was confirmed by differential expression analysis for myeloid cells and macrophages comparing FMRP-KO with WT tumors (table S3). Of these, three proinflammatory chemokines known to recruit CD8 T cells (*45*–*50*) stood out: *Ccl5*, *Cxcl9*, and *Cxcl10*, which were broadly up-regulated in myeloid cells and macrophages from FMRP-KO tumors (Fig. 7, B to D, and table S3). The principal sources of these three cytokines in the TME were monocytes and macrophages, with the exception of *Ccl5*, which is also highly expressed in lymphocytes (fig. S11, A to C). Given that these chemokines are known to be up-regulated in myeloid cells by infiltrating CD8 T cell–induced interferon-γ (IFN-γ) secretion, we first considered whether expression of these genes in TAMs and myeloid cells in FMRP-KO tumors was regulated directly by FMRP-KO cancer cells or indirectly by IFN-γ–secreting T cells. To this end, WT and FMRP-KO cancer cells were inoculated into immunodeficient mice, which enabled the evaluation of chemokine expression in the absence of an adaptive immune system. Analysis of FACS-purified TAMs revealed that CCL5 and CXCL10 were expressed in FMRP-KO tumors in the absence of T cells (Fig. 7E), consistent with direct, inductive communication from macrophages to lymphocytes. The same analysis was then performed in immunocompetent mice, which revealed that the presence of an adaptive immune system increased the number of TAMs expressing CCL5 and induced CXCL9 expression in TAMs, whereas CXCL10 was unchanged compared with the TAMs from tumors in the immunodeficient mice (Fig. 7F). The expression of these three proinflammatory chemokines was also documented in CT26 tumors (fig. S11D).

Next, aiming to further substantiate the inferred mechanism of T cell–dependent amplification of TAMs expressing these proinflammatory chemokines, PDAC KO2 tumors were treated with anti–IFN-γ antibody, which, when compared with sham-treated PDAC-KO tumors, resulted in enhanced tumor weight and reduced CD8 T cell abundance (Fig. 7G) as well as lower CTL activity reflected in a FACS analysis that assessed granzyme B (GZMB) and tumor necrosis factor–α (TNF-α) expression (fig. S11E). Consistent with the expression pattern in immunodeficient mice (Fig. 7F), the neutralizing antibody to IFN-γ reduced CCL5 and CXCL9 expression in tumor-infiltrating macrophages, whereas CXCL10 expression was evidently independent of IFN-γ–expressing CD8 T cells (fig. S11F).

Having established the direct and indirect up-regulation of three proinflammatory chemokines in TAMs populating FMRP-KO tumors, we assessed their functional importance for the T cell–dependent tumor immunity by pharmacologically inhibiting their receptors. Thus, mice bearing FMRP-KO2 tumors were treated with a small molecule inhibitor—maraviroc (*51*)—of the receptor for CCL5 (fig. S11G) or with a blocking antibody for CXCR3 (*52*), the receptor for CXCL9 and CXCL10 (fig. S11H). The two inhibitors, most notably in combination (Fig. 7H), reduced the recruitment of activated CD8 T cells and rescued the otherwise impaired growth rate of the FMRP-KO tumors compared with vehicle or immunoglobulin G (IgG)–treated controls (Fig. 7H and fig. S11, I and J). The three chemokines CCL5, CXCL9, and CXCL10 were also modestly expressed in TAMs from WT tumors detected in the FACS analysis (Fig. 7F and fig. S11D), although mRNA levels were very low (Fig. 7B). However, treatment of mice bearing WT tumors with the inhibitors of the CCL5, CXCL9, and CXCL10 receptors had no effect on either tumor growth or survival (fig. S11, K and L), which suggests that their expression is insufficient to elicit T cell recruitment in the highly immunosuppressive FMRP-WT TME, consistent with the low levels of infiltrating T cells (Fig. 3, A and B and Fig. 4D) and the absence of antitumoral T cell immunity (Fig. 3, I and J, and fig. S6G). As such, these three chemokines, up-regulated in the absence of FMRP, are established as indirectly regulated secondary components of the FMRP-KO immunostimulatory phenotype, operating as downstream effectors of the recruitment and antitumoral efficacy of T cells; notably, the chemokine signaling through the CCR5 and CXCR3 receptors is functionally independent and complimentary, such that their combined activity is required for robust CD8 T cell inflammation and consequent antitumor immunity.

Collectively, the results presented above elucidate the distinctive immune phenotypes consequent to FMRP expression versus its absence in tumors, as summarized in Fig. 7I. FMRP-expressing cancer cells induce immunosuppressive T_{regs} and M2-like TAMs through the expression of IL-33 and of PROS1 plus exosomes, respectively, producing a TME essentially devoid of T cells. The down-regulation of these three immunosuppressive regulatory factors in tumors with genetic KOs or KDs of FMRP in cancer cells, in concert with up-regulation of immunostimulatory CCL7, leads to the recruitment and activation of T cells that are responsible for the impaired tumor growth and improved survival of mice bearing tumors lacking this regulator of translation and transcription.

Associating FMRP's cancer network with prognosis of human cancers

Having uncovered and functionally characterized the unprecedented activity of FMRP through functional manipulations in mice, we sought to assess the translational applicability, to human cancers, of the realization that FMRP can limit the recruitment of CD8 T cells into tumors and thereby facilitate tumor growth. To begin, we assessed the predictive value of *FMR1* mRNA expression and failed to associate comparatively elevated levels with overall

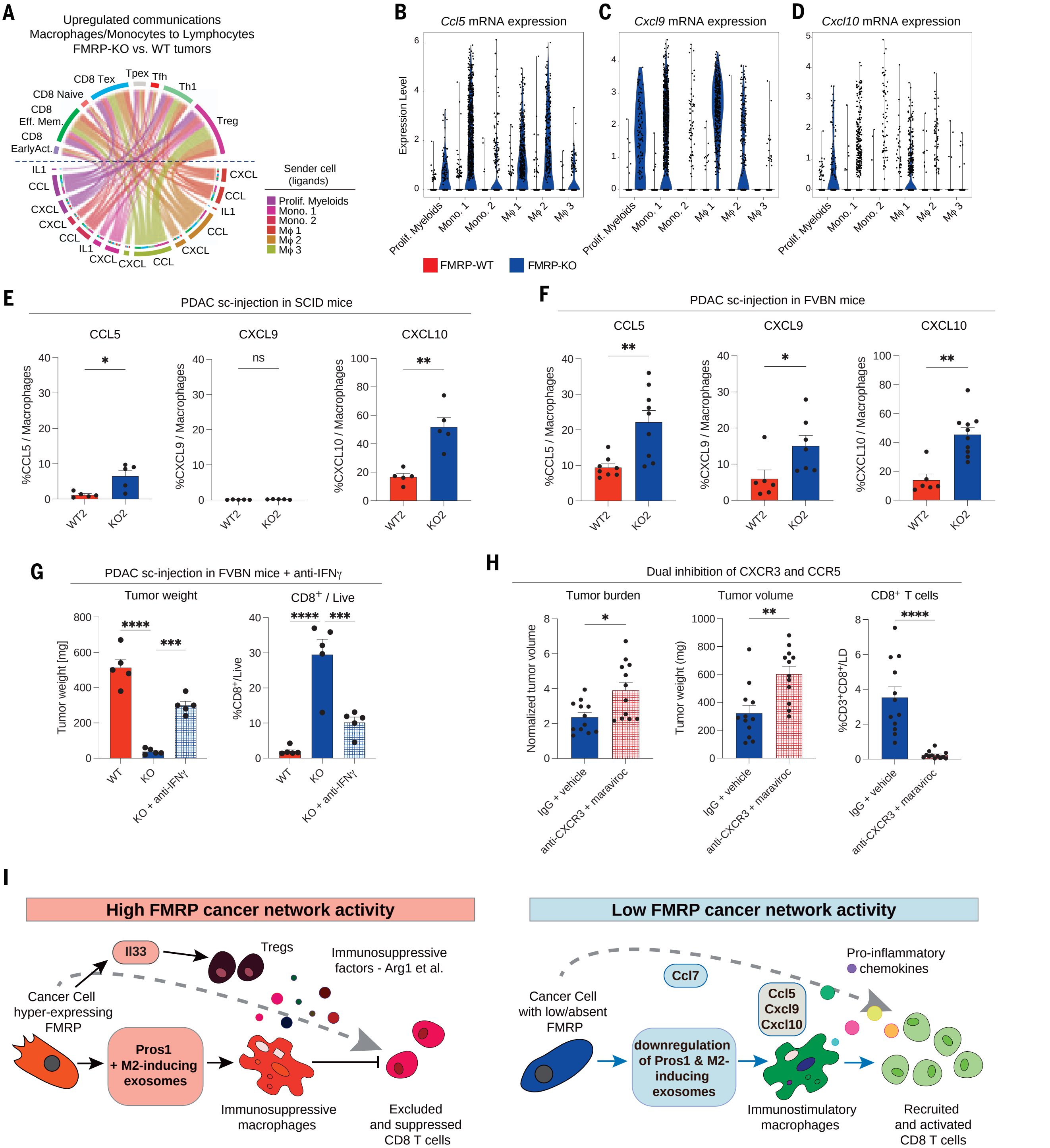


Fig. 7. Proinflammatory chemokines directly and indirectly up-regulated in macrophages of FMRP-KO tumors orchestrate inflammation by T cells that mediate tumor immunity. (**A**) Inferred intercellular communication networks in FMRP-KO tumors from myeloid cells as senders of ligands to lymphocytes as receivers expressing cognate receptors in FMRP-KO tumors, revealing up-regulation of CCL and CXCL pathways in FMRP-KO myeloid cells (see also table S1 for specifics). (**B** to **D**) Violin plot showing the distribution of mRNA expression for differentially up-regulated ligands in myeloid cells comparing FMRP-KO versus WT tumors: *Ccl5* (B), *Cxcl9* (C), and *Cxcl10* (D). (**E** and **F**) FACS analysis of CCL5 (left), CXCL9 (middle), and CXCL10 (right) ligands expression in the overarching macrophage population—identified as being CD45$^+$/CD11b$^+$/Ly6C$^-$/Ly6G$^-$—within WT versus FMRP-KO tumors derived from inoculation of PDAC cancer cells growing as tumors in SCID mice (E) or FVBN mice (F). Statistical significance was assessed with the Mann-Whitney test. (E) N = 5 mice for all arms. (F) For CCL5, N = 8 for the WT cohort and N = 9 mice for the KO2 cohort; for CCL9, N = 6 for WT and 7 for KO cohorts; and for CXCL10, N = 6 for WT and N = 10 for KO cohorts; one independent experiment. Data are represented as means ± SEMs. (**G**) Quantification of tumor burden (left) and CD8 T cell infiltration (right) in PDAC tumors at day 14 postinoculation (s.c.) of WT and FMRP-KO cancer cells, as well as FMRP-KO cancer cells treated with an IFN-γ neutralizing antibody. One-way ANOVA was used for multiple group comparisons. N = 5 mice for each group; one independent experiment. Data are

presented as means ± SEMs. (**H**) Quantification of tumor burden (left), tumor volume (middle), and FACS analysis of CD8 T cell infiltration (right) of FMRP-KO tumors in mice treated with the combination the CCR5 inhibitor maraviroc and anti-CXCR3 antibody. *P* values were assessed with the Mann-Whitney test. *N* = 12 mice for each group; one independent experiment. Data are presented as means ± SEMs. (**I**) Schematic representation of FMRP-mediated immunosuppressive mechanisms in tumors. Cancer cells hyperexpressing FMRP modulate T_{regs} and immunosuppressive macrophages cells through the up-regulation of IL-33 and PROS1 and exosomes, respectively, which prevents appreciable inflammation by CD8 and CD4 T lymphocytes (left). By contrast, cancer cells lacking FMRP recruit CD8 and CD4 T lymphocytes and promote their proliferation through up-regulation in cancer cells of CCL7 and down-regulation of IL-33, along with down-regulation of PROS1 and exosomes, whose absence leads to indirect activation of proinflammatory cytokines (CCL5, CXCL9, and CXCL10) in reprogrammed immunostimulatory macrophages; these profound differences in the TME collectively trigger antitumor immunity, leading to impaired tumor growth and improved survival. **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001.

survival (OS), with disease-free survival (DFS), or with reduced CD8 T cell infiltration (estimated by a CD8 T cell signature; see Materials and methods) in a pan-cancer analysis of 31 human tumor datasets from TCGA (fig. S12, A to C). This negative result is potentially consistent with the knowledge that FMRP is not only a regulator of translation and transcription but is also itself posttranscriptionally regulated in neurons, such that the mRNA levels, as well as protein levels, do not necessarily correlate linearly with its downstream translational network activity (*53*, *54*).

Given that the level of *FMR1* mRNA was not prognostically informative, we developed an overarching gene signature that is indicative of FMRP's regulation of a downstream transcriptional network in cancer cells (henceforth called the FMRP cancer signature), which reflects genes in the network that are directly and indirectly controlled by FMRP at the transcriptional level. The FMRP cancer signature was generated by combining the differentially expressed genes in two subsignatures, derived by comparing FMRP-WT versus FMRP-KO PDAC cell lines in culture [FMRP cancer cell (CC) subsignature] and by comparing FMRP-WT versus FMRP-KO tumors (FMRP tumor subsignature) (fig. S12D, table S4, and Materials and methods). Functional enrichment analysis linked the FMRP cancer signature with multiple pathways (fig. S12E and table S4) relating to inflammation and immune responses as well as epithelial-mesenchymal transition (EMT) and other tumor phenotypes potentially reflective of FMRP's previously reported capability to stimulate invasion and metastasis (*15*–*17*). Notably, the correlation of differentially expressed genes with the apoptosis pathway is only evident in the FMRP tumor subsignature, derived from scRNA-seq data of WT versus FMRP-KO tumors, consistent with ongoing apoptosis induced by the observed CTL inflammation and killing, and not by intrinsic differences in the WT versus FMRP-KO cancer cells. To assess the specificity of the FMRP CC subsignature and to ensure that it is directly linked to FMRP expression (and not, for example, to adaptations occurring during outgrowth of the single-cell *Fmr1-KO* clones), *Fmr1* was knocked down in a PDAC cancer cell line, using siRNA transfection (fig. S12F), and RNA-seq analysis was performed on samples collected after 48 hours. Upon overlaying the FMRP CC subsignature onto the siRNA FMRP-KD transcriptome profile, we observed a significant down-regulation of the FMRP CC signature score in FMRP-KD cancer cells (fig. S12G), consistent with its veracity as a reporter for FMRP's regulatory network, which we therefore applied in the analyses below.

In notable contrast to the nonassociation of *FMR1* mRNA itself, the FMRP cancer signature revealed a statistically significant association with OS and DFS across the same cohort of 31 cancer datasets, such that patients with a higher FMRP cancer signature score have worse OS and DFS (Fig. 8A and fig. S12H). Moreover, a high FMRP cancer signature score demonstrated a statistically significant anticorrelation with a CD8 T cell signature that is diagnostic of CTL abundance in human tumors (Fig. 8B). Notably, the FMRP cancer signature score exhibits similar distributions across four clinical tumor stages reflective of malignant progression of different cancer types (fig. S12I), which suggests that the observed associations with prognosis are not driven by FMRP's reported roles in stimulating invasion and metastasis.

To substantiate these results, we further investigated the correlation between patient survival and the FMRP CC subsignature, which only involves differentially expressed genes in bulk RNA-seq analysis of the WT versus KO PDAC cell lines in culture and therefore evaluates FMRP's downstream transcriptional network without the indirect effects of the TME (fig. S12D and table S4). Consistently, the FMRP CC subsignature also revealed a significant correlation with patient OS and DFS across the 31 cancer types (Fig. 8C and fig. S12J) as well as a significant anticorrelation with CD8 T cell infiltration in tumor samples (Fig. 8D), which was again independent of tumor stage (fig. S12, K and L). These results further suggest that FMRP transcriptional network activity per se is limiting the recruitment of CD8 T cells into human tumors and influencing cancer patients' prognosis.

The association of the FMRP cancer signature with reduced OS and lower CD8 T cell infiltration is further illustrated for three human tumor types, namely endometrial carcinoma, melanoma, and head and neck squamous cell carcinoma, which have particularly provocative discriminations between tumors with high versus low FMRP signature scores (Fig. 8, E to G). Again, no meaningful associations with tumor aggressiveness (estimated by tumor stage; fig. S12, M to O) were observed. Congruently, similar inverse associations with CD8 infiltration scores were observed with the FMRP CC subsignature in these three tumor types (fig. S12, P to R), substantiating the implication that FMRP transcriptional network activity is modulating the immune response in human tumors.

Furthermore, the FMRP cancer signature showed a significant correlation with progression-free survival (PFS) in colorectal cancer (Fig. 8H, left panel), wherein lymphocytic reaction is an important prognostic factor (*55*). Notably, this association exists for microsatellite stable (MSS) but not for microsatellite instable (MSI) colorectal patients (Fig. 8H, middle and right panels). Additionally, the FMRP CC subsignature showed a similarly significant association with PFS in colorectal cancer, again only for MSS tumors (fig. S13A). To assess the effect of tumor grade on the observed association, the samples were separated into lower-grade (stages I to III) and high-grade (stage IV) tumors. A similar trend was observed for the association of the FMRP cancer signature with PFS in lower-grade MSS (but not MSI) tumor samples (fig. S13B). The high-grade colorectal cancers represented in the available dataset comprise only MSS tumor samples, for which the FMRP activity signature again showed a significant correlation with PFS (fig. S13C). Notably, MSI patients with high mutational burden often demonstrate better response to immunotherapy and tend to have better prognosis (*56*). The less-immunogenic MSS tumors, however, can evade the immune response and metastasize in >80% of cases with advanced colorectal cancer (*57*). The association of the FMRP signature scores with the prognosis of MSS patients suggests that high FMRP transcriptional network activity may contribute to the poor responses in this and other tumor types largely refractory to immune checkpoint inhibition (see below).

Breast cancers have been histologically characterized as immune deserts, largely devoid of inflammation by CD8 T cells (*58*). Despite this overall poor immunogenicity, human breast tumor samples with a low FMRP signature score are characterized by a significantly higher CD8 T cell infiltration score in a TCGA cohort of

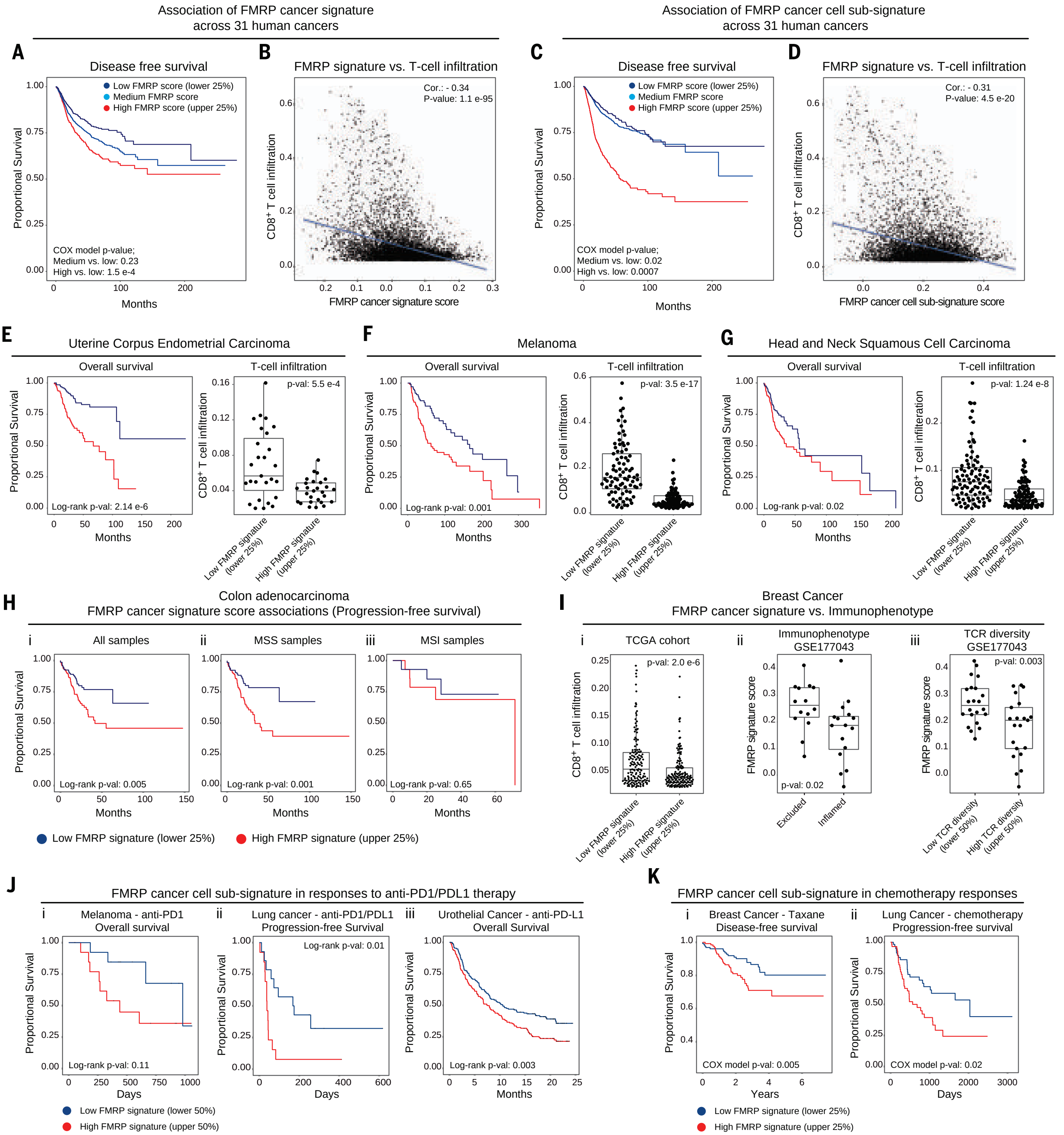


Fig. 8. FMRP transcriptional network signatures are negatively associated with both CD8 T cell infiltration and more favorable prognosis in multiple human tumors. (**A**) Anticorrelation of the FMRP cancer signature score with DFS in the TCGA human pan-cancer dataset. Low FMRP activity, quantile 1, blue curve; medium FMRP activity, quantiles 2 and 3, light blue curve; high FMRP activity, quantile 4, red curve. The COX model was used, considering the tumor type as a covariate to estimate the significance of correlation. (**B**) Anticorrelation of the FMRP cancer signature score with a CD8 T cell infiltration signature, estimated by the xCell package, in human pan-cancers. Linear regression model with tumor type as a covariate was used to estimate the significance of correlation. (**C**) Anticorrelation of the FMRP CC subsignature score with DFS in the TCGA human pan-cancer dataset, as in (A). (**D**) Anticorrelation of the FMRP CC subsignature score with a CD8 T cell infiltration signature, as in (B). Only up-regulated genes in FMRP activity CC subsignature were used in deriving the signature score in this analysis. (**E** to **G**) Anticorrelation of the FMRP cancer signature score with PFS (left) and CD8 T cell infiltration signature (right; boxplots showing median, lower, and upper quantiles) in endometrial carcinoma (E), melanoma (F), and head and neck squamous cell carcinoma (G). Log-rank test was used for survival analyses, and Wilcoxon two-tailed test was used for the CD8 T cell association analyses. (**H**) Anticorrelation of the FMRP cancer signature score with PFS for human colorectal cancer. Significant correlation was found in all colorectal samples (i) and for MSS colorectal samples (ii), but

with a notable lack of correlation for MSI samples (iii); see also fig. S12, A to C. (**I**) FMRP cancer signature score in human breast cancer, showing boxplots with median, lower, and upper quantiles for the signature scores. (i) Comparison of CD8 T cell infiltration score in high versus low FMRP cancer signature–scored tumor samples. (ii) Comparison FMRP cancer signature scores in immune-excluded versus inflamed breast cancer tumors (cohort: GSE177043). (iii) Comparison FMRP cancer signature scores in low– versus high–TCR diversity breast cancer tumors (cohort: GSE177043). Statistics by Wilcoxon two-tailed test. (**J**) Correlation of the FMRP CC subsignature score (fig. S13K) with poor prognosis for different cancer patients in response to immune checkpoint inhibitor therapy. (i) Association of a high FMRP CC subsignature score with comparatively poor OS in melanoma patients treated with anti–PD-1. (ii) Association of a high FMRP CC subsignature score with worse PFS of lung cancer patients treated with anti–PD-1 or anti–PD-L1. (iii) Association of a high FMRP CC subsignature score with poorer OS in urothelial cancer patients treated with anti–PD-L1. (**K**) Correlation of a high FMRP CC subsignature score with comparatively poor prognosis for cancer patients in response to different forms of chemotherapy. (i) Association of a high FMRP CC subsignature score with reduced DFS of breast cancer patients treated with taxane. (ii) Association of a high FMRP CC subsignature score with poorer DFS of lung cancer patients treated with cisplatin, carboplatin, or paclitaxel.

breast cancer patients (Fig. 8I, panel i), which is again independent of tumor stage (fig. S13D). To further investigate the relevance of the FMRP signature, we leveraged a recently published cohort of 43 primary breast cancer patients (*59*) that includes transcriptional profiling of tumor samples with matched clinical, immunohistochemistry and multiplexed immunofluorescence data collectively describing a multiparametric immunophenotype. Consistent with the results presented above, a high FMRP cancer signature score was associated with low numbers of tumor-infiltrating lymphocytes (immune-excluded immunophenotype; Fig. 8I, panel ii) and low TCR diversity in tumors (Fig. 8I, panel iii). Likewise, the FMRP CC subsignature score showed a significant correlation with this immune-excluded phenotype in the tumor samples of this dataset (fig. S13E). Although the transcriptomics data are a rich resource for estimating CD8 T cell infiltration in tumor samples, to further validate our results, we sought to evaluate possible associations at the protein level. To this end, a TMA cohort of breast cancer patients from Sweden Cancerome Analysis Network - Breast (SCAN-B) was used to assess FMRP protein expression level alongside CD8 T cell abundance through immunostaining for both (fig. S13F). Using the paired transcriptomics data for the TMA samples, we first revealed a significantly positive correlation between *FMR1* mRNA and FMRP protein levels (fig. S13G). Next, the TMA samples were categorized as high or low for CD8 infiltration, based on the immunostaining quantification. Congruent with the previous results, no meaningful associations—in particular anticorrelations—were observed between CD8 T cell abundance and FMRP mRNA or protein expression levels (fig. S13, H and I). In marked contrast, the FMRP cancer signature score for the SCAN-B TMA samples again revealed a significant inverse correlation with CD8 T cell abundance (fig. S13J), which further supports the role of FMRP transcriptional network activity in modulating the immune response in human breast cancer and substantiates the relevance and potential utility of the FMRP cancer network signatures in the clinical setting.

Next, we assessed datasets of melanoma, lung, and urothelial cancer patients that were treated with and scored for responses to immune checkpoint inhibitors and biopsied for transcriptome analysis. Notably, tumors with a high FMRP CC signature score responded poorly or failed to respond to anti–PD-1 or anti–PD-L1 therapies (fig. S13K), and a high FMRP CC subsignature score correlated significantly with poorer prognosis of treated patients (Fig. 8J). It is notable that the FMRP CC subsignature was used in this analysis, which suggests that the observed associations were solely based on intrinsic FMRP transcriptional cancer network activity, without effects of the TME involving indirectly induced immune response genes.

Finally, recognizing that optimal efficacy for certain forms of chemotherapy is now appreciated to require an adaptive immune response (*60*–*63*), we assessed the FMRP CC subsignature score in several cohorts of cancer patients treated with chemotherapy. Consistent with FMRP's demonstrable role in programing an immunoevasive TME, a high FMRP cancer cell–derived transcriptional cancer network signature had a statistically significant association with poorer prognosis in breast and lung cancer patients in the context of treatment with different forms of immune-dependent chemotherapy (Fig. 8K). These associations were again independent of tumor stage (fig. S13L), emphasizing the role of the FMRP regulatory network as an immunosuppressor throughout tumor progression. In this case again, the FMRP CC subsignature was used, whereby the FMRP regulatory network in cultured cancer cells was queried for its collective diagnostic effects in the treated human tumors without the potentially confounding effects of the immune-inflamed TME.

Discussion

Our findings suggest that FMRP limits the recruitment and expansion of CD8 and CD4 T cells, which allows solid tumors to evade immune destruction. This capability was revealed by experimental tumor models based on engineered gene KOs of the *Fmr1* gene (encoding the FMRP protein) in a series of cancer cell lines of different origins, all of which expressed endogenous FMRP at a high level. The aberrant up-regulation of FMRP was evident at varying frequencies in a variety of human cancers and genetically engineered mouse models, including many tumor types that are infrequently responsive to immune checkpoint immunotherapy. A gene signature diagnostic of the cancer transcriptional network regulated by FMRP was associated with worse DFS and OS and reduced inflammation by CD8 T cells in a pan-cancer analysis and in selected human tumor types, consistent with the hypothesis that FMRP is a cancer cell–intrinsic immunosuppressor that contributes to immune evasion and the pathogenesis of human cancer. An intriguing question is why should a protein involved in synaptic transmission have this capability? One possibility is that FMRP serves to raise the bar to protect the brain against precocious induction of T cell inflammation, as suggested in a recent study in a rat model of neuroinflammation (*13*).

Notably, we and others have previously implicated FMRP as a promoter of invasion and metastasis (*15*–*17*), which has not been a focus in this study. However, our bioinformatic studies did not significantly associate high signature scores for the FMRP cancer regulatory network with tumor grade or stage in multiple forms of human cancer (fig. S12, I and K to O).

FMRP is an RNA binding protein that has diverse functions in neurons (*8*, *64*), of which the most prominent is the translational regulation of hundreds of only partially overlapping proteins in different cell types (*7*, *65*–*67*). Additionally, FMRP binds to and potentially modulates the functions of multiple cellular proteins involved in diverse functions (*66*), and it also directly and indirectly regulates mRNA levels of hundreds of genes through transcription, mRNA splicing, and mRNA stability (*7*, *8*). Although the constellation of FMRP's translational regulatory targets in cancer cells has not yet been elaborated, we have shown by transcriptome profiling that FMRP significantly modulates the mRNA levels of more than a hundred genes in WT versus FMRP-KO cancer cells and indirectly affects a similar number of genes in the cell types populating the TME, which supports the proposition that it is a key regulator of transcriptional and translational networks in cancer cells.

Our investigation suggests that FMRP creates a barrier to immune recruitment that is multifactorial, as are the immunostimulatory

effects in its absence. As described above, we have begun to dissect the functional effectors of these dichotomous immune phenotypes. Accordingly, analysis of the TME in FMRP-WT versus FMRP-KO tumors by scRNA-seq and cellular annotation (Fig. 4) has revealed significant alterations of the adaptive and innate immune compartments and has led to the identification and functional validation of molecular determinants involved in the immune evasion versus tumor immunity orchestrated by FMRP up-regulation versus its absence, respectively (Fig. 7I).

The immunoevasive TME in FMRP-expressing tumors involves both T_{regs} and immunosuppressive (M2-like) macrophages (and potentially additional classes of myeloid cells), each elicited by factors secreted by FMRP-expressing cancer cells. The T_{regs} were induced by IL-33, which we have implicated through CLIP-seq as a direct target of FMRP translational control. The importance of IL-33 was confirmed in functional studies, wherein it was overexpressed in FMRP-KO cancer cells. Concordantly, immunosuppressive M2-like macrophages were induced by PROS1 secretion, implicated by CLIP-seq as another direct target of FMRP activity, again functionally validated in coculture assays and through overexpression in FMRP-KO cancer cells. Concomitantly, cancer cell–derived exosomes have been implicated as yet another primary mechanistic effector of FMRP expression in cancer cells, whereby these exosomes demonstrably induced immunoinhibitory gene expression in macrophages as part of their programming to become immunosuppressive.

By contrast, the inflammatory TME composed of activated CD8 and CD4 T cells that develops in FMRP-deficient tumors has several mechanistic components originating in cancer cells lacking FMRP expression. IL-33 was down-regulated, which reduces the abundance of immunosuppressive T_{regs}, as did the up-regulation of otherwise FMRP-repressed CCL7, which concomitantly contributed to the recruitment and proliferative capability of CD8 and CD4 T cells and the T cell–dependent impairment of tumor growth. Additionally, PROS1 was down-regulated in FMRP-deficient cancer cells, which contributed to the attenuation of immunosuppressive macrophages, as did the quantitative reduction and qualitative modification of cancer cell–secreted exosomes. Although it is conceivable that unidentified M1-inducing factors are up-regulated in FMRP-deficient cancer cells, these two effectors are demonstrably instrumental in inducing immunosuppressive M2-like macrophages, thereby allowing, consequent to their down-regulation in FMRP-deficient cancer cells, the development of immunostimulatory (M1-like) macrophages. The net result is the appearance in tumors populated by FMRP-deficient cancer cells of immunostimulatory macrophages expressing the proinflammatory chemokines CCL5, CXCL9, and CXCL10, which indirectly contributed to the recruitment of CD8 T cells that in turn amplified chemokine expression and T cell inflammation through the secretion of IFN-γ.

The molecular mechanisms governed by FMRP and its absence that we have functionally implicated in the immunosuppressed versus inflamed tumors are summarized in Fig. 7I. It is evident, e.g., from the scRNA-seq analysis, that FMRP—much like in neurons—is a key regulator of a multiplicity of genes, operating both directly (and indirectly) in cancer cells and indirectly in cells of the TME. As such, we expect that other components remain to be identified and functionally elucidated.

Finally, consistent with the knowledge base about FMRP's complex posttranscriptional and posttranslational regulation in neurons (53), we have not been able to associate the levels of *FMR1* mRNA (or in the case of TNBC, the levels of FMRP protein) with prognosis in human cancers (fig. S12, A to C, and fig. S13, F to I), which implicates analogous regulatory mechanisms operative in cancer cells that warrant future investigation. This disconnect motivated the development of gene signatures diagnostic of FMRP's regulation of a cancer network (fig. S12D and table S4), which have revealed FMRP's downstream transcriptional network as a negative prognostic factor for many forms of human cancer (Fig. 8; fig. S12, H and J; and fig. S13, A to C and K). Evidence that immune evasion is an important prognostic component, congruent with the mechanistic studies presented herein, comes from the demonstrations that high FMRP transcriptional network activity was associated with comparatively poor responses to immune checkpoint inhibitors in several patient cohorts for which transcriptome datasets are available (Fig. 8J and fig. S13K) and with poor responses to certain chemotherapies (Fig. 8K and fig. S13L) for which the activity of the adaptive immune system has been implicated as an instrumental part of therapeutic efficacy (61). As such, it is possible that the described gene signatures for FMRP cancer network expression (or protein sets derived therefrom) could have utility in selecting patients more likely to respond to such immune-dependent therapies. Overall, the widespread expression of FMRP in solid tumors, concomitant with induction of its cancer regulatory network, constitutes a previously unappreciated mechanism whereby tumors evade immune destruction.

Materials and methods

Patient tumor TMA analysis

The human PDAC and colon cancer TMAs were constructed at the Institute of Pathology, University of Bern. (Patient samples were collected in accordance with the Swiss Federal Human Research Act (HFG 2014). All tissues were retrieved from the archives of the Institute of Pathology, University of Bern, Switzerland. Tissue blocks were used for the construction of the TMAs. Permission to use the patient material and accompanying patient-related data for this study were approved by the Ethics committee of the Canton of Bern and were collected in accordance with the 2014 Human Research Act/Law (KEK#200/2014). Samples were coded after they were retrieved from the archives of the Institute of Pathology, University of Bern, and matched to the corresponding data. The details of the construction of those TMAs and have been described previously (68, 69).

The HCC TMA was constructed at the Institute of Pathology, University of Bern, Switzerland. Tissue blocks were collected for the construction of the TMA. Permission to use patient material for this study was approved by the Ethics committee of the Canton of Bern (KEK#07/10/13). A total of 119 patients diagnosed with HCC were included from 1991 to 2008. Multiple—up to seven—core biopsies (0.6 mm) of HCC tumors were included for every patient. Clinical data except for the diagnosis of HCC are not available.

The prostate cancer brain metastasis TMA was constructed as part of the larger prostate cancer brain metastases cohort (PCBM cohort; $N = 51$) (unpublished; currently in peer-review). The TMA contains 204 core biopsies (each 1 mm in diameter) from prostate cancer brain metastases (PCBMs) as well as—in some cases—matched primary tumors (PTs) and benign tissue (BT), comprising 108, 66, and 30 core biopsies, respectively. Multiple core biopsies from both PCBM and matched primaries covering intratumoral heterogeneous areas were included in the TMA, identified by a previous review protocol based on histomorphological features and immunohistochemical expression. Additionally, core biopsies from BT were incorporated when available. All analyses were carried out in accordance with protocols approved by the Ethical Committee Bern (project ID: 2019 – 00328). One of the 5 TMA blocks constructed was used for the FMRP analysis. This block contains tissue from 15 different patients including 6 patients harboring all tissue types, 6 patients with PCBM and BT, and 3 patients with PT and BT. Note that this is a cohort of a special subtype of prostate cancers that develop brain metastasis, which may not reflect the subtypes of prostate cancers without brain metastasis.

The human TNBC TMA was constructed at the University of Lund, in accordance with appropriate ethical approvals, as described previously (70). Sections on glass slides from this TMA were sent to the Institute of Pathology in Bern for immunohistochemical analysis of

FMRP expression. Normal breast tissues were collected from women undergoing reduction mammaplasties with no history of breast cancer, as previously described (71), and tissue sections provided by the Brisken laboratory (EPFL, Lausanne) were used for immunostaining to detect FMRP expression.

Generation of CRISPR-edited Fmr1-null cancer cell lines

Mouse PDAC 4361.12 cells, a single clone–derived cell line from the mouse PDAC cell line 4361 (15) were cultured in Dulbecco's minimum essential medium (DMEM) with 10% fetal bovine serum (FBS) (Gibco, catalog no. 10270-106) with 100 U penicillin / 0.1 mg/ml streptomycin (Thermo Fisher Scientific, catalog no. 15140122). The *Fmr1* gene was deleted in cancer cells using the CRISPR-Cas9 system, to produce so-called FMRP-KO cells. Cultured cancer cells were transiently cotransfected with Cas9 and single-guide RNA (sgRNA) expression plasmids (72) and selected in blasticidin (10 ug/ml; Thermo Fisher Scientific, catalog no. A1113903) for 5 days. This transient CRISPR strategy for the deletion of the *Fmr1* gene avoids potential unspecific effects consequent to the stable integration of Cas9/sgRNA into the genome. The guide sequences used for *Fmr1* were 5′-GTGGAAGTGCGGGGCTCCAA-3′ and 5′- GAGCTGGTGGTGGAAGTGCG-3′. Single cells were distributed into 96-well culture plates without blasticidin. *Fmr1* KO clones were identified by immunoblotting of protein lysates for the lack of FMRP protein using two FMRP antibodies (table S5). The obtained *Fmr1*-deleted (FMRP-KO) cells remained sensitive to blasticidin, indicating the CRISPR-Cas9 system had been only transiently expressed in those cells. Two independent FMRP-KO clones and two control WT clones (transfected with a Cas9 vector and the empty sgRNA vector) were analyzed as described in the results section.

Similarly, the *Fmr1* gene encoding FMRP was deleted in mouse CT26 colon cancer cells, 4T1 breast cancer cells, and B16-OVA melanoma cells. In brief, cultured cancer cells were transiently cotransfected with Cas9 and an *sgFmr1* expression plasmid or the empty sgRNA vector and selected in blasticidin for 5 days. Single cells were then distributed into 96-well plates, and the FMRP KO clones were validated by immunoblotting. One WT and one or two FMRP-KO clones were analyzed as indicated for the described experiments.

CT26 and 4T1 cells were cultured in RPI-1640, whereas B16-OVA cells were cultured in DMEM, all with 10% FBS with 100 U penicillin / 0.1 mg/ml streptomycin. All cell lines were tested as mycoplasma negative.

For the mixed inoculation experiment, to fluorescently label the WT2 PDAC cells with RFP and FMRP-KO2 cells with GFP, cells were infected with an RFP or GFP lentivirus (GEG-tech, France, lenti-rfpluc, lenti-gfpluc) according to the manufacturer's instructions.

Engineering the Ccl7-KO and FMRP-KO PDAC cells

The *Ccl7* gene was deleted in mouse PDAC 4361.12 FMRP KO2 cells using the Alt-R CRISPR-Cas9 system (Integrated DNA Technologies), to produce so-called FMRP/Ccl7 DKO cells. Cultured cancer cells were cotransfected with Cas9 protein and CRISPR RNA (crRNA) targeting *Ccl7*, and transactivating CRISPR RNA (tracrRNA) labeled with ATTO 550 (tracrRNA-ATTO 550; Integrated DNA Technologies, catalog no. 1075927) using electroporation (Nepagene, NEPA21 Super Electroporator). Single cells were distributed into 96-well culture plates after cells were recovered from the electroporation chamber. *Ccl7* KO clones were identified by enzyme-linked immunosorbent assay (ELISA) analysis of supernatants for the lack of CCL7 secretion. The crRNA sequences used for *Ccl7* are 5′-AAGCGTGGCAGAGATCCTCA-3′, 5′-CGTGG-CAGAGATCCTCATGG-3′, and 5′-AAGCAGCG-CCTATGAGCAGC-3′.

Engineering the expression vectors to restore expression of Pros1 and Il-33

The overexpression plasmid pLenti-C-Myc-DDK-P2A-Puro (OriGene, catalog no. PS100092), as well as mouse *Pros1* cDNA (OriGene, catalog no. MR224303) and *Il-33* cDNA (OriGene, catalog no. MR203528), were purchased from OriGene. The insertion of the *Pros1* and *Il-33* cDNAs into the overexpression plasmid used the Precision Shuttling system from OriGene. The construction of lentiviruses using the overexpression plasmid followed the lentiviral packaging protocol provided by OriGene. FMRP KO2 cells were infected with lentivirus produced with mock and *Pros1* or *Il-33* rescue vectors were selected using Puromycin (2 ug/ml; Thermo Fisher Scientific, catalog no. A1113803).

siRNA transfection

Parental cells were seeded at 100,000 cells per well into a 6-well plate and transfected the next day with 20 μM *siRNA* (5 μl per well, 100 pmol final) using RNAiMAX lipofectamine (5 μl per well) in OptiMEM Reduced Serum Media (Thermo Fisher Scientific, catalog no. 13778075). The transfected cells were then used for in various bioassays, including the CD4/CD8 T coculture assay, the M1 macrophage coculture assay, and the CTL cell killing assay, 48 hours posttransfection. MISSION *siRNA* Universal Negative Control no. 1 and no. 2 (Sigma-Aldrich, catalog nos. SIC001 and SIC002) were used as the *siCtrl1* and *siCtrl2* constructs, respectively. The sequences used for siRNAs targeting *Fmr1* (Stealth siRNA) were GAGUU-CAAGGCAGCUUGCCUCAAGA (*siFmr1*) and GAGCUAGUUCUAGACCACCACCAAA (*siFmr2*), purchased from Thermo Fisher Scientific. For siRNA targeting Myc, the sequence was CGAGAACAGUUGAAACACA, purchased from Sigma-Aldrich.

Cell proliferation and viability assays

Cell proliferation was measured using an MTT cell proliferation kit (Roche, catalog no. 11 465 007 001). Briefly, 1×10^4 cells were plated in triplicate in 96-well plates. 72 hours later, 10 μl of MTT labeling reagent was added to each well and then incubated for 4 hours at 37°C, followed by the addition of 100 μl MTT solubilization reagent overnight. Absorbance was measured at 59 nm on a plate reader (Tecan Safire). For the CellTiter-Glo Luminescent Cell Viability Assay (Promega, catalog no. G7570), 1×10^4 cells were seeded in one well of a 96-well plate. 24 hours later, cell viability was measured according to the manufacturer's protocol.

Cell apoptosis assay

1.5×10^5 cells were seeded into each well of a 6-well tissue culture plate. 24 hours later, Annexin-V^+/7-AAD^- apoptotic cells were measured by flow cytometry (Attune NxT, Thermo Fisher Scientific), according to the manufacturer's protocol (BioLegend, catalog no. 640922).

Cell migration assay

5000 cells in 50 μl basal medium were seeded into the top inserts of a 96-well Boyden chamber (Corning, catalog no. 3374); 200 μl of complete medium was added to each bottom well. 18 hours later, nonmigratory cells growing on the top membrane of the inserts were removed with a cotton swab infused with 70% ethanol, and migratory cells underneath the membrane were fixed, stained with crystal violet, and counted.

Reverse transcription polymerase chain reaction (RT-PCR)

RNA was isolated with the miRNeasy Mini Kit (Qiagen, catalog no. 217084) or RNeasy Plus Micro Kit (Qiagen, catalog no. 74034). A total of 500 ng of RNA was reverse transcribed into cDNA using the PrimeScript RT Master Mix (Takara, catalog no. RR036A). cDNA levels were quantified using the SYBR green method in a QuantStudio 6 Flex Qpcr system in a 384-well format. 18*S* ribosomal RNA was used to equilibrate sample loading. The primers used in these studies are listed in table S6.

Protein (Western) blotting

Cells were lysed in RIPA Lysis and Extraction Buffer (Thermo Fisher Scientific, catalog no. 89900). Protein concentration was assessed using the BCA Protein Assay Kit (Thermo Fisher Scientific, catalog no. 23225), or the Bradford 1X dye reagent (Bio-Rad). Samples were resolved using Mini-Protean precast gels and transferred to PDVF or NC membranes. Membranes were blocked with 5% BSA/TBST for 1 hour at room temperature and incubated overnight at 4°C with antibodies recognizing ARG1, IL-10, Hsp90,

FMRP, PROS1, Alix, CD81, GAPDH, Myc, and Ovabumin (table S5). Membranes were then probed with HRP-linked secondary antibodies and developed with WesternBright Sirius (Advansta) or the PierceTM ECL Protein Blotting substrate (Thermo Fisher Scientific) using Fusion FX7.

ELISA for CCL7

1×10^6 PDAC WT2, WT3, FMRP-KO2, and FMRP-KO8 cells were seeded into a 100 mm tissue culture plate with DMEM complete medium on day 1. On day 3, medium was removed and replaced with 6 ml of serum-free DMEM medium. On day 5, medium was collected into 15 ml tubes, and centrifuged at 2000 rpm in centrifuge at 4°C for 10 min. Supernatants were collected and diluted at 1:2 ratio. For the ELISA analysis of CCL7 in WT2, KO2, and KO2-CCL7 KO cells, 0.1×10^6 cells were seeded into one well of 6-well plate in DMEM complete medium, and supernatants were collected 48 hours later. The secretion of CCL7 in the supernatants was analyzed using the mouse MCP3 ELISA Kit (CCL7) (Abcam, catalog no. ab205571), following the manufacturer's protocol.

Animal studies

All experiments using animals were performed in accordance with protocols approved by the local animal experimentation committee of the Canton de Vaud (license no. 3214 and 3214.x). FVB/N, Balb/c, C57Bl/6, NSG, or SCID/beige mice were used at 8 weeks of age. For the lung metastasis assay, 2×10^5 mouse PDAC WT or FMRP-KO cells suspended in 200 µl phosphate-buffered saline (PBS) were injected into the lateral tail vein of the mice. For the PT growth assay, 5×10^5 cells suspended in 100 µl PBS were injected s.c. For the cell mixing injection experiment, PDAC WT-RFP and FMRP-KO-GFP cells engineered to express RFP or Turbo GFP protein (as described above) were mixed at different ratios; in total 5×10^5 cells suspended in 100 µl PBS were injected s.c. Mice were monitored twice per week and euthanized at the days indicated in the figure legends.

For the orthotopic tumor growth assay in the pancreas, 10-week-old FVBn mice were injected with 200,000 cells in a total volume of 50 µl into the parenchyma of the pancreas (in close proximity to the spleen). Mice were euthanized 2 weeks later and tumors collected for analysis.

For pharmacological trials using function-blocking antibodies (table S5), mice were treated intraperitoneally with 250 µg of a given antibody per mouse twice per week. The CCR5 inhibitor Maraviroc (Selleckchem, catalog no. S2003) was prepared in PBS and administered daily by intraperitoneal injection at 10 mg/kg. All treatments were initially started on the same day as the s.c. cancer cell inoculation, and continued for 2 weeks, except for the intervention trial of anti–PD-1, which was started when the tumors reached a volume of 100 mm^3.

Analysis of tumor-infiltrating lymphocytes by flow cytometry

For intracellular cytokine staining, mice were injected 6 hours before euthanasia with 250 µg Brefaldin A (Sigma-Aldrich, catalog no. B7651). To obtain single-cell suspensions, harvested tumors were chopped into fragments, digested in a mix of DNase I (144 U/mL, Roche, catalog no. 11284932001), Dispase II (0.85 U/mL, Roche, catalog no. 04 942 078 001) and Collagenase A (0.33 U/mL, Roche, catalog no. 10 103 578 001), and passed through a 70-µm cell strainer. Cell suspensions were then blocked with CD16/32 (clone 93, BioLegend) and labeled with live/dead reagent (The LIVE/DEAD Fixable Red Dead Cell Stain Kit, Thermo Fisher Scientific, catalog no. L34971, or LIVE/DEAD Fixable Blue Dead Cell Stain Kit, for UV excitation, Thermo Fisher Scientific, catalog no. L12105). Surface staining was performed for 20 min on ice.

To detect biotinylated antibodies, samples were incubated with fluorophore-conjugated streptavidin. Cells were then fixed and permeabilized with the Foxp3/Transcription Factor Staining Buffer Set Kit (Invitrogen, catalog no. 00-5523-00), and intracellular staining was performed in a Perm/Wash buffer overnight at 4°C. Samples were variously analyzed on Gallios or Cytoflex (Beckman Coulter), LSRII Fortessa (BD), or Attune NxT (Thermo Fisher Scientific) flow cytometers. Data analysis was performed using FlowJo software. The antibodies used in this analysis are shown in table S5.

Immunohistochemical and immunofluorescent staining

Harvested mouse tissues were fixed in 4% paraformaldehyde (PFA) overnight, embedded in paraffin, and sectioned using a microtome (Leica). Antigen retrieval was performed in a citrate buffer (pH = 6.0) in a water bath at 95°C for 20 min, or in a tris-EDTA buffer (pH = 8.0) in a water bath at 95°C for 10 min. Primary antibodies were incubated at 4°C overnight.

For immunohistochemical (IHC) staining, secondary antibodies (ImmPRESS HRP reagent kit, anti-rabbit, catalog no. MP-7401, and anti-rat, catalog no. MP-7444) were incubated at room temperature for 45 min and finally visualized with the peroxidase substrate DAB (Sigma-Aldrich, catalog no. D5637-1G) for the same amount of time (maximum 10 min) at room temperature. The stained tissue sections were counterstained with Meyer's hematoxylin.

For immunofluorescent (IF) staining, secondary antibodies (Alexa Fluor 488, 568, and 647; Thermo Fisher Scientific, catalog nos. A-11006, A-11077, and A-21247) were incubated at room temperature for 45 min.

For the IF staining of fresh tumor slides from the mixing injection experiment, the WT-RFP and KO2-GFP cells were mixed at 1:1 ratio; in total 5×10^5 cells suspended in 100 µl PBS were injected s.c. On day 7, mice were anesthetized and perfused with 20 ml PBS and 20 ml 4% PFA. Tumors were collected and fixed in 4% PFA for 4 more hours, and then embedded in 2% agarose in PBS and sectioned onto 100 µm slides with a Leica Biosystems vibratome. Slides were incubated with primary CD8 antibody (table S5) at 4°C for 2 days, and then secondary antibody (goat anti-rat Alexa Fluor 647; Thermo Fisher Scientific, catalog no. A-21247) was applied overnight.

Images were variously acquired with a Leica DM5500B microscope, a Zeiss LSM 700 upright confocal microscopes, or an Olympus VS120 whole-slide scanner, and analyzed with Image J. The antibodies used are shown in table S5.

Ex vivo T cell coculture assays

CD8 and CD4 T cells were isolated from a spleen cell suspension following the EasySep Mouse CD8 or CD4 T cell isolation kits (StemCell Technologies, catalog no. 19853 or 19851, respectively). Cells were then stained with CFSE (BioLegend, catalog no. 422701) or Cell tracer violet (Thermo Fisher Scientific, catalog no. C34571) at 1 µM for 6 min and resuspended in the RPMI media containing 10% FBS, 1% PS, NEAA (Thermo Fisher Scientific, catalog no. 11140050) and β-mercaptoethanol (Thermo Fisher Scientific, catalog no. 21985023). T cells were then activated with the CD3/CD28 Dynabeads (Thermo Fisher Scientific, catalog no. 11456D) and plated at 200,000 cells per well in a 96-well plate. Cancer cells were then added at a 1:1 ratio.

For the CD11b cocultures, tumors were harvested and processed as described in the flow cytometry section. CD11b cells were then isolated from the cell suspension following the EasySep Mouse CD11b Positive Selection Kit II (StemCell, catalog no. 18970). After isolation, cells were plated in a 1:1 ratio with T cells. 72 hours later, the $CSFE^{-low}$ CD4 or CD8 T cells were counted by flow cytometry (Gallios, Beckman Coulter).

In vitro T cell killing assay

3×10^3 B16-OVA WT and FMRP-KO cells were seeded into a 96-well plate for 48 hours (or 24 hours). OT1 CD8+ T cells were isolated from the splenocytes of OVA-specific T cell receptor transgenic OT-1 mice, and then activated by OVA peptide plus IL-2 and IL-7 (10 ng/ml; BioLegend, catalog nos. 575406 and 577806) for 5 days following a published protocol (*73*). 3×10^3 preactivated OT1 CD8 T cells were then added to cancer cell culture at different ratios, with IL-2 (10 ng/ml) included to maintain their activity. 24 hours later, the viable cancer cells were counted by flow cytometry (Attune NxT, Thermo Fisher Scientific, or Cytoflex, Beckman Coulter).

BMDM isolation

Bone marrow extraction was performed as previously described (74). Cells were plated in RPMI with 10% FBS, 1% Pen/Strep and 50 ng/mL mouse macrophage colony-stimulating factor (BioLegend, catalog no. BLG-576406-100 uG). On day 7, BMDMs were polarized using cytokines and cancer cells. For M1-like polarization, 200 U/mL IFN-γ (Proteintech, catalog no. 315-05-20UG) and 100 ng/mL LPS (Sigma-Aldrich, catalog no. L4516-1MG) were used. To induce M2-like polarization, 20 ng/mL Il-4 (Sigma-Aldrich, catalog no. I1020-5UG) was used. For the coculture experiments, PDAC WT or FMRP KO cancer cells were seeded at 1×10^5 cells/mL into a 0.4-μm insert a day before the polarization experiment. The next day, inserts containing cancer cells were transferred to a 6-well plate seeded with M1- or M2-polarized BMDMs. Cells were harvested 48 hours after polarization for further analyses. For FACS analyses, 1.5×10^5 cancer cells in total were plated directly onto polarized BMDMs for 48 hours and then collected and stained as described in the flow cytometry section of the Materials and methods. For the experiments involving Pros1, macrophages were treated with 1 μg/mL mouse recombinant Pros1 protein (R&D Systems, catalog no. 9740-PS). The exosome cocultures followed a previously described protocol (75, 76).

Preparation of EVs

Isolation of EVs was performed as previously reported (75, 76). Briefly, PDAC cells were plated in 15-cm cell culture dishes. When cells reached 50% confluency, complete medium was replaced with 5% EV-depleted FBS medium. The EVs were collected after ~2 to 3 days using sequential centrifugation, as described. Quantification was performed with the BCA Protein Assay Kit.

Analyzing ENCODE ChIP-seq data for TFs binding to the FMR1 promoter region

All human MAX and MYC ChIP-seq data from ENCODE database were downloaded. The following two types of files are used: (i) *.bed* files were used to find TFs' binding location on FMR1's promoter, for which the peak calling algorithm has already been applied, and (ii). *bigWig* files were used for representing the data (fig. S3).

We have also used the Ensembl database for selecting FMR1's promoter region: Chromosome X: 147,911,000-147,915,201, whose stable ID is ENSR00000249258. The result for the significant peak is presented in table S7.

TCR sequencing

WT2 or FMRP-KO2 PDAC cancer cells (5×10^5) were s.c. injected into the right flank of FVBn mice. 14 days later, PDAC tumors were dissected, and genomic DNA extracted (using Qiagen DNeasy Blood & Tissue Kit, catalog no. 69504). Immunosequencing of the Mouse TCRβ complementary determining region 3 (CDR3) was performed using the ImmunoSEQ Assay (Adaptive Biotechnologies, Seattle, WA). In brief, the mouse CDR3 regions of T cells from the whole-tumor genomic DNAs were amplified through bias-controlled multiplex PCRs using primers specific to TCR-Vβ and TCR-Jβ segments and then subjected to high-throughput sequencing. The resulting raw sequences were filtered to identify and quantitate the absolute abundance of each individual TCR rearrangement in each sample. $N = 3$ mice for each group.

Mutational load and neoantigen prediction analysis in cell lines

WT and FMRP KO PDAC cells (WT2, WT3, KO2, and KO8) were cultured in DMEM with 10% FBS (Gibco) and 100 U penicillin / 0.1 mg/ml streptomycin (Gibco) for 30, 60, and 150 days; the cells were spit when 90% confluent and passaged every 3 days. Genomic DNAs was extracted from the cells at these different time points (Qiagen, catalog no. 69504), and whole-exome sequencing was performed using an Illumina HiSeq 4000.

Exome sequences were mapped to the mouse genome using BWA (v.0.7.15), and variants were called using FreeBayes (v.1.1.0) against mm10. Germline variants were filtered out using the mouse strain FVBN/J germline variations [single-nucleotide polymorphisms (SNPs) and indels] stored in the Mouse Genome Project (www.sanger.ac.uk/science/data/mouse-genomes-project).

Variant annotation was performed with VariantAnnotation R package. Amino acid coding changes were predicted for the nonsynonymous variants; the SIFT (sorting intolerant from tolerant) and PolyPhen (polymorphism phenotyping) algorithms provided predictions of how severely the coding changes might be affecting protein function. The SIFT method uses sequence homology and the physical properties of amino acids to make predictions about protein function. PolyPhen uses sequence-based features and structural information characterizing the substitution to make predictions about the structure and function of the protein. The remaining variants were remapped to the FVBN/J_v1 genome for neoantigen prediction. MuPeXi was used to predict all mutated peptides (neopeptides) of lengths 8 to 11, using FVBN/J-specific peptide and cDNA references.

scRNA-seq analysis

WT (WT2, WT3) and FMRP-KO (KO2, KO8) PDAC cancer cells (5×10^5) were separately injected s.c. into the right flank of FVBN mice. 14 days later, PDAC tumors were dissected and digested by collagenase II, IV (0.25%), and DNase I (0.05%) in DMEM/F12 medium for 30 min at 37°C. Cells were counted, and 8×10^3 primary cancer cells for each tumor were loaded into the 10X Genomics Chromium platform. Samples were processed following the manufacturer's protocol with Single Cell 3′ v2 reagent and then sequenced using an Illumina NextSeq sequencer. Raw data were mapped to mouse genome mm10-3.0.0 with cellranger-3.1.0. Downstream analyses were performed in R (v.3.6.1) using Seurat (v.3.2.2) and scran (v.1.14.6) following the guidelines given in http://bioconductor.org/books/release/OSCA/. We used standard Seurat settings for normalization, principal components analysis (PCA), t-distributed stochastic neighbor embedding (tSNE) analysis, and clustering. For differential expression analysis, the Wilcoxon rank sum test was applied, using 1.5 for the average fold change in expression level, and 0.05 as the cutoff for adjusted P value.

scRNA-seq cluster annotation

To annotate different clusters of scRNA-seq analysis, we used cell type–specific gene signature lists derived from multiple references (33, 77, 78), applying the singscore R package (79) to derive signature scores for each gene-list. Subsequently, scRNA-seq clusters were annotated according to the highest gene-list signature scores (fig. S7). Cell type–specific gene signature lists were used as follows—cancer cells: Krt18, Krt8, and Krt19; B cells: Ms4a1, Bank1, Cd79b, Fcer2a, Pax5, Cd79a, Tcl1, Stap1, Fcrl5, Pou2af1, Ptprc, and Cd19; lymphocytes: Cd5, Ubash3a, Il7r, Itk, Cd28, Themis, Bcl11b, Cd3d, Cd3e, and Cd3g; DCs: Itgax, Cst3, and Cd74; migDC: Ccr7, Ccl22, and Fscn1; cDC1: Itgae, Clec9a, Ccl17, and Xcr1; pDC: Bst2, Ly6d, Siglech, and Tcf4; neutrophils: Pglyrp1, Csf3r, Cyp4f18, and Itgam; myeloid cells: Kit, Cd34, Ly6a, Kcnip3, Med21, Sgms2, Tfec, Itgam, and Ly6g; monocytes: Ms4a6c, Ms4a6b, Ms4a6d, and Ly6c2; macrophages: Adgre1, Cd68, Csf1r, Fcgr1, Cd14, C1qb, C1qa, Flt3, Cx3cr1, and Vcam1; CAFs: Col1a1, Col1a2, Pdpn, and Dcn; myCAFs: Tagln, Thy1, Col12a1, and Thbs2; iCAFs: Clec3b, Col14a1, Has1, and Il6; pericytes: Acta2, Higd1b, Rgs5, and Pdgfrb; and endothelial cells: Pecam1.

After annotating each cluster, to ensure that the clusters within each cell type had similar distributions of the total number of features (reads), we filtered out the clusters that had a low number of reads (average < 1000 read) such that their distributions were significantly lower (3 standard deviations lower) than the rest of the clusters with the same annotations. Therefore, one cancer cell cluster and two myeloid cell clusters with very low number of reads were excluded from the subsequent analysis.

Lymphocyte analysis, ProjecTILs

All clusters annotated as lymphocytes were subselected, and further analyzed for lymphocyte subtype annotations. To this end, we used the ProjecTILs pipeline, implemented in an R

package (*33*), which is an algorithm that accurately overlays scRNA-seq data onto reference T cell atlases, and characterizes different cell states and clusters.

To assess natural killer (NK) cells, which are not incorporated into ProjecTILs, we used scGate (*80*), an algorithm that identifies scRNA-seq populations based on both positive and negative markers. We applied scGate onto the whole lymphocyte subpopulation using NCAM1^{pos}, KLRD1^{pos}, CD3D^{neg} as the virtual gating strategy, and annotated the positive cells as the NK cell population.

Cell-cell communication analysis in scRNA-seq data

To infer the outgoing communication patterns between different cell types in the scRNA-seq data, we applied the CellChat pipeline (*34*), implemented in the R package (https://github.com/sqjin/CellChat), following the pipeline provided by the developers. Accordingly, the cell-cell communication patterns between cell types within each condition, namely WT and FMRP-KO tumors, were separately inferred, using the "computeCommunProb" function. Then, the respective ligand-receptor networks from the two conditions were merged using "mergeCellChat" function. Subsequently, to identify the up- and down-regulated signaling ligand-receptor pairs between WT and FMRP-KO tumors, we used the "netVisual_bubble" function with the default parameters in the package. To further filter the identified interactions, the differentially expressed genes (comparing WT with FMRP-KO tumors) were extracted for the source cell type (the cell population expressing the ligands), and then only the interactions for which the ligands are within this list of differentially expressed genes were selected and plotted as the up- or down-regulated signaling ligand-receptor pairs between two cell populations.

mRNA sequencing analysis

We performed RNA-seq analysis for two independent KO clones and WT clones. For one clone of each condition (KO2 and WT2), we performed the RNA-seq in two separate baches. As for PDAC WT2 cells with siCTLR and siFMR1 transfection, we performed the RNA-seq in one batch. RNA-seq libraries were generated from double-stranded cDNA derived from 10 ng total RNA with the NuGEN Ovation RNA-Seq System V2 (NuGEN Technologies, San Carlos, CA). 100 ng of the cDNA was fragmented to 350 bp using Covaris S2 (Covaris, Woburn, MA). Sequencing libraries were prepared from the fragmented cDNA with the Illumina TruSeq Nano DNA Library Prep Kit (Illumina, San Diego, CA). Cluster generation was performed with the libraries using the Illumina TruSeq SR Cluster Kit v4 reagents and sequenced on the Illumina HiSeq 2500 with TruSeq SBS Kit v4 reagents. Sequencing data were processed using the Illumina Pipeline Software version 1.82. Adapter sequences and low-quality ends were removed with cutadapt,1.4.2, trimming for TrueSeq and polyA sequences. Reads were aligned to mouse genome build mm10 using HISAT2 aligner. Normalization was performed using DESeq2 pipeline, by normalizing with the size factor. This method is implemented in the R Bioconductor package DESeq2. Differential expression analysis was performed by DESeq pipeline (DESeq2 R package). The cutoffs for gene selection were adjusted $P < 0.05$ and fold change > 1.5.

Development of gene signatures for FMRP pathway activity

To identify a gene signature reflecting FMRP downstream transcriptional network, we developed three different signature lists, as follows (fig. S11D):

1) Differentially expressed genes (fold change > 1.5, adjusted $P < 0.05$) were identified, comparing WT and FMRP-KO cell lines. Because we had two baches of FMRP-KO2 clones, only differentially expressed genes in both batches were selected, and then the union of these genes with the differentially expressed genes in the FMRP-KO8 clone constituted the list of genes in FMRP CC network subsignature.

2) Differentially expressed genes (fold change > 1.5, adjusted $P < 0.05$) between WT and FMRP-KO tumors from the scRNA-seq data were identified and constitute the list of genes in FMRP tumor network subsignature. In this analysis, we included all the cancer cells detected using Krt8 and Krt18 marker genes, regardless of the total number of features.

3) The union of these two gene lists constitutes the main FMRP cancer network signature.

HALLMARK terms enrichment analysis

HALLMARK enrichment analysis was performed using Molecular Signatures Database (MSigDB), available through an online service. False discovery rate (FDR)–adjusted $P < 0.05$, retrieving a maximum of 20 terms.

Gene signature enrichment analysis

Single-sample gene set enrichment analysis (ssGSEA) (*81*) was used to calculate an expression score for gene expression signatures. The method was implemented in the R package GSVA.

Correlation analysis

To evaluate the correlation between two continuous variables, "cor" function was used in R to develop the Pearson correlation coefficient. To estimate the significance of the correlation, we used the "lm" function in R for linear modeling and retrieved FDR-adjusted *P* values.

Survival analysis

Kaplan-Meier survival analysis was used to assess the relationship of the signature scores and FMR1 expression with OS. We applied the large-sample Chi-square test (log-rank test) to determine the associations between predictor variables and to obtain adjusted hazard-ratios. These analyses were performed with the R package "survival."

Quantification and statistical analysis

All statistical analyses described above were performed either with R, or with Prism 7 (GraphPad Software).

REFERENCES AND NOTES

1. J. Gill, V. Prasad, A reality check of the accelerated approval of immune-checkpoint inhibitors. *Nat. Rev. Clin. Oncol.* **16**, 656–658 (2019). doi: 10.1038/s41571-019-0260-y; pmid: 31383994
2. R. W. Jenkins, D. A. Barbie, K. T. Flaherty, Mechanisms of resistance to immune checkpoint inhibitors. *Br. J. Cancer* **118**, 9–16 (2018). doi: 10.1038/bjc.2017.434; pmid: 29319049
3. A. J. Schoenfeld, M. D. Hellmann, Acquired Resistance to Immune Checkpoint Inhibitors. *Cancer Cell* **37**, 443–455 (2020). doi: 10.1016/j.ccell.2020.03.017; pmid: 32289269
4. M. S. Sidorov, B. D. Auerbach, M. F. Bear, Fragile X mental retardation protein and synaptic plasticity. *Mol. Brain* **6**, 15 (2013). doi: 10.1186/1756-6606-6-15; pmid: 23566911
5. J. C. Darnell *et al*., FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011). doi: 10.1016/j.cell.2011.06.013; pmid: 21784246
6. Y. Feng *et al*., FMRP associates with polyribosomes as an mRNP, and the I304N mutation of severe fragile X syndrome abolishes this association. *Mol. Cell* **1**, 109–118 (1997). doi: 10.1016/S1097-2765(00)80012-X; pmid: 9659908
7. M. Li *et al*., Identification of FMR1-regulated molecular networks in human neurodevelopment. *Genome Res.* **30**, 361–374 (2020). doi: 10.1101/gr.251405.119; pmid: 32179589
8. J. D. Richter, X. Zhao, The molecular biology of FMRP: New insights into fragile X syndrome. *Nat. Rev. Neurosci.* **22**, 209–222 (2021). doi: 10.1038/s41583-021-00432-0; pmid: 33608673
9. C. Malecki, B. D. Hambly, R. W. Jeremy, E. N. Robertson, The RNA-binding fragile-X mental retardation protein and its role beyond the brain. *Biophys. Rev.* **12**, 903–916 (2020). doi: 10.1007/s12551-020-00730-4; pmid: 32654068
10. E. J. Mientjes *et al*., The generation of a conditional *Fmr1* knock out mouse model to study Fmrp function in vivo. *Neurobiol. Dis.* **21**, 549–555 (2006). doi: 10.1016/j.nbd.2005.08.019; pmid: 16257225
11. The Dutch-Belgian Fragile X Consortium, *Fmr1* knockout mice: A model to study fragile X mental retardation. *Cell* **78**, 23–33 (1994). doi: 10.1016/0092-8674(94)90569-X; pmid: 8033209
12. E. Donnard, H. Shu, M. Garber, Single cell transcriptomics reveals dysregulated cellular and molecular networks in a fragile X syndrome model. *PLOS Genet.* **18**, e1010221 (2022). doi: 10.1371/journal.pgen.1010221; pmid: 35675353
13. P.-Y. Nie *et al*., miR-142 downregulation alleviates rat PTSD-like behaviors, reduces the level of inflammatory cytokine expression and apoptosis in hippocampus, and upregulates the expression of fragile X mental retardation protein. *J. Neuroinflammation* **18**, 17 (2021). doi: 10.1186/s12974-020-02064-0; pmid: 33407653
14. D. Prilutsky *et al*., Gene expression analysis in *Fmr1*KO mice identifies an immunological signature in brain tissue and mGluR5-related signaling in primary neuronal cultures. *Mol. Autism* **6**, 66 (2015). doi: 10.1186/s13229-015-0061-9; pmid: 26697163
15. L. Li *et al*., GKAP Acts as a Genetic Modulator of NMDAR Signaling to Govern Invasive Tumor Growth. *Cancer Cell* **33**, 736–751.e5 (2018). doi: 10.1016/j.ccell.2018.02.011; pmid: 29606348
16. R. Lucá *et al*., The fragile X protein binds mRNAs involved in cancer progression and modulates metastasis formation. *EMBO Mol. Med.* **5**, 1523–1536 (2013). doi: 10.1002/emmm.201302847; pmid: 24092663
17. F. Zalfa *et al*., The fragile X mental retardation protein regulates tumor invasiveness-related pathways in melanoma cells. *Cell Death Dis.* **8**, e3169 (2017). doi: 10.1038/cddis.2017.521; pmid: 29144507
18. A. Di Grazia *et al*., The Fragile X Mental Retardation Protein Regulates RIPK1 and Colorectal Cancer Resistance to Necroptosis. *Cell. Mol. Gastroenterol. Hepatol.* **11**, 639–658 (2021). doi: 10.1016/j.jcmgh.2020.10.009; pmid: 33091622

19. G. Pedini *et al*., FMRP modulates the Wnt signalling pathway in glioblastoma. *Cell Death Dis.* **13**, 719 (2022). doi: 10.1038/s41419-022-05019-w; pmid: 35982038
20. M. Uhlén *et al*., A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteomics* **4**, 1920–1932 (2005). doi: 10.1074/mcp.M500279-MCP200; pmid: 16127175
21. F. Sanchez-Vega *et al*., Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **173**, 321–337.e10 (2018). doi: 10.1016/j.cell.2018.03.035; pmid: 29625050
22. Y. Luo *et al*., New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.* **48**, D882–D889 (2020). doi: 10.1093/nar/gkz1062; pmid: 31713622
23. S. Pelengaris, M. Khan, G. Evan, c-MYC: More than just a matter of life and death. *Nat. Rev. Cancer* **2**, 764–776 (2002). doi: 10.1038/nrc904; pmid: 12360279
24. N. M. Sodir *et al*., MYC Instructs and Maintains Pancreatic Adenocarcinoma Phenotype. *Cancer Discov.* **10**, 588–607 (2020). doi: 10.1158/2159-8290.CD-19-0435; pmid: 31941709
25. J. C. Castle *et al*., Immunomic, genomic and transcriptomic characterization of CT26 colorectal carcinoma. *BMC Genomics* **15**, 190 (2014). doi: 10.1186/1471-2164-15-190; pmid: 24621249
26. J. R. Brahmer *et al*., Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N. Engl. J. Med.* **366**, 2455–2465 (2012). doi: 10.1056/NEJMoa1200694; pmid: 22658128
27. N. Pu, W. Lou, J. Yu, PD-1 immunotherapy in pancreatic cancer: Current status. *J. Pancreatol.* **2**, 6–10 (2019). doi: 10.1097/JP9.0000000000000010
28. T. R. Medler *et al*., Complement C5a Fosters Squamous Carcinogenesis and Limits T Cell Response to Chemotherapy. *Cancer Cell* **34**, 561–578.e6 (2018). doi: 10.1016/j.ccell.2018.09.003; pmid: 30300579
29. A. Y. Huang *et al*., The immunodominant major histocompatibility complex class I-restricted antigen of a murine colon tumor derives from an endogenous retroviral gene product. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9730–9735 (1996). doi: 10.1073/pnas.93.18.9730; pmid: 8790399
30. T. Moroishi *et al*., The Hippo Pathway Kinases LATS1/2 Suppress Cancer Immunity. *Cell* **167**, 1525–1539.e17 (2016). doi: 10.1016/j.cell.2016.11.005; pmid: 27912060
31. L. M. Kranz *et al*., Systemic RNA delivery to dendritic cells exploits antiviral defence for cancer immunotherapy. *Nature* **534**, 396–401 (2016). doi: 10.1038/nature18300; pmid: 27281205
32. X. Zhang *et al*., CellMarker: A manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.* **47**, D721–D728 (2019). doi: 10.1093/nar/gky900; pmid: 30289549
33. M. Andreatta *et al*., Interpretation of T cell states from single-cell transcriptomics data using reference atlases. *Nat. Commun.* **12**, 2965 (2021). doi: 10.1038/s41467-021-23324-4; pmid: 34017005
34. S. Jin *et al*., Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021). doi: 10.1038/s41467-021-21246-9; pmid: 33597522
35. T. Maurin *et al*., HITS-CLIP in various brain areas reveals new targets and new modalities of RNA binding by fragile X mental retardation protein. *Nucleic Acids Res.* **46**, 6344–6355 (2018). doi: 10.1093/nar/gky267; pmid: 29668986
36. Y. Liu, Y. Cai, L. Liu, Y. Wu, X. Xiong, Crucial biological functions of CCL7 in cancer. *PeerJ* **6**, e4928 (2018). doi: 10.7717/peerj.4928; pmid: 29915688
37. A. Hatzioannou *et al*., An intrinsic role of IL-33 in T_{reg} cell–mediated tumor immunoevasion. *Nat. Immunol.* **21**, 75–85 (2020). doi: 10.1038/s41590-019-0555-2; pmid: 31844326
38. D. G. DeNardo, B. Ruffell, Macrophages as regulators of tumour immunity and immunotherapy. *Nat. Rev. Immunol.* **19**, 369–382 (2019). doi: 10.1038/s41577-019-0127-6; pmid: 30718830
39. S. Yan, G. Wan, Tumor-associated macrophages in immunotherapy. *FEBS J.* **288**, 6174–6186 (2021). doi: 10.1111/febs.15726; pmid: 33492779
40. E. Ubil *et al*., Tumor-secreted Pros1 inhibits macrophage M1 polarization to reduce antitumor immune response. *J. Clin. Invest.* **128**, 2356–2369 (2018). doi: 10.1172/JCI97354; pmid: 29708510
41. M. S. Pols, J. Klumperman, Trafficking and function of the tetraspanin CD63. *Exp. Cell Res.* **315**, 1584–1592 (2009). doi: 10.1016/j.yexcr.2008.09.020; pmid: 18930046
42. A. L. Wozniak *et al*., The RNA binding protein FMR1 controls selective exosomal miRNA cargo loading during inflammation. *J. Cell Biol.* **219**, e201912074 (2020). doi: 10.1083/jcb.201912074; pmid: 32970791
43. Q. Chen *et al*., Exosome-Mediated Crosstalk Between Tumor and Tumor-Associated Macrophages. *Front. Mol. Biosci.* **8**, 764222 (2021). doi: 10.3389/fmolb.2021.764222; pmid: 34722637
44. Z. Jiang *et al*., Cancer derived exosomes induce macrophages immunosuppressive polarization to promote bladder cancer progression. *Cell Commun. Signal.* **19**, 93 (2021). doi: 10.1186/s12964-021-00768-1; pmid: 34521440
45. M. T. Chow *et al*., Intratumoral Activity of the CXCR3 Chemokine System Is Required for the Efficacy of Anti-PD-1 Therapy. *Immunity* **50**, 1498–1512.e5 (2019). doi: 10.1016/j.immuni.2019.04.010; pmid: 31097342
46. I. G. House *et al*., Macrophage-Derived CXCL9 and CXCL10 Are Required for Antitumor Immune Responses Following Immune Checkpoint Blockade. *Clin. Cancer Res.* **26**, 487–504 (2020). doi: 10.1158/1078-0432.CCR-19-1868; pmid: 31636098
47. N. Karin, CXCR3 Ligands in Cancer and Autoimmunity, Chemoattraction of Effector T Cells, and Beyond. *Front. Immunol.* **11**, 976 (2020). doi: 10.3389/fimmu.2020.00976; pmid: 32547545
48. J. Liu *et al*., Local production of the chemokines CCL5 and CXCL10 attracts $CD8^+$ T lymphocytes into esophageal squamous cell carcinoma. *Oncotarget* **6**, 24978–24989 (2015). doi: 10.18632/oncotarget.4617; pmid: 26317795
49. P. M. Marcovecchio, G. Thomas, S. Salek-Ardakani, CXCL9-expressing tumor-associated macrophages: New players in the fight against cancer. *J. Immunother. Cancer* **9**, e002045 (2021). doi: 10.1136/jitc-2020-002045; pmid: 33637602
50. T. J. Zumwalt, M. Arnold, A. Goel, C. R. Boland, Active secretion of CXCL10 and CCL5 from colorectal cancer microenvironments associates with GranzymeB+ CD8+ T-cell infiltration. *Oncotarget* **6**, 2981–2991 (2015). doi: 10.18632/oncotarget.3205; pmid: 25671296
51. G. G. Xu, J. Guo, Y. Wu, Chemokine receptor CCR5 antagonist maraviroc: Medicinal chemistry and clinical applications. *Curr. Top. Med. Chem.* **14**, 1504–1514 (2014). doi: 10.2174/1568026614666140827143745; pmid: 25159165
52. R. Uppaluri *et al*., Prolongation of cardiac and islet allograft survival by a blocking hamster anti-mouse CXCR3 monoclonal antibody. *Transplantation* **86**, 137–147 (2008). doi: 10.1097/TP.0b013e31817b8e4b; pmid: 18622291
53. M. Prieto, A. Folci, S. Martin, Post-translational modifications of the Fragile X Mental Retardation Protein in neuronal function and dysfunction. *Mol. Psychiatry* **25**, 1688–1703 (2020). doi: 10.1038/s41380-019-0629-4; pmid: 31822816
54. C. M. Rodriguez *et al*., A native function for RAN translation and CGG repeats in regulating fragile X protein synthesis. *Nat. Neurosci.* **23**, 386–397 (2020). doi: 10.1038/s41593-020-0590-1; pmid: 32066985
55. S. Ogino *et al*., Lymphocytic reaction to colorectal cancer is associated with longer survival, independent of lymph node count, microsatellite instability, and CpG island methylator phenotype. *Clin. Cancer Res.* **15**, 6412–6420 (2009). doi: 10.1158/1078-0432.CCR-09-1438; pmid: 19825961
56. S. Kang *et al*., The significance of microsatellite instability in colorectal cancer after controlling for clinicopathological factors. *Medicine* **97**, e0019 (2018). doi: 10.1097/MD.0000000000010019; pmid: 29489646
57. G. Golshani, Y. Zhang, Advances in immunotherapy for colorectal cancer: A review. *Therap. Adv. Gastroenterol.* **13**, 1756284820917527 (2020). doi: 10.1177/1756284820917527; pmid: 32536977
58. S. Bayraktar, S. Batoo, S. Okuno, S. Glück, Immunotherapy in breast cancer. *J. Carcinog.* **18**, 2 (2019). doi: 10.4103/jcar.JCar_2_19; pmid: 31160888
59. D. Hammerl *et al*., Spatial immunophenotypes predict response to anti-PD1 treatment and capture distinct paths of T cell evasion in triple negative breast cancer. *Nat. Commun.* **12**, 5668 (2021). doi: 10.1038/s41467-021-25962-0; pmid: 34580291
60. L. Galluzzi, J. Humeau, A. Buqué, L. Zitvogel, G. Kroemer, Immunostimulation with chemotherapy in the era of immune checkpoint inhibitors. *Nat. Rev. Clin. Oncol.* **17**, 725–741 (2020). doi: 10.1038/s41571-020-0413-z; pmid: 32760014
61. J. W. Opzoomer, D. Sosnowska, J. E. Anstee, J. F. Spicer, J. N. Arnold, Cytotoxic Chemotherapy as an Immune Stimulus: A Molecular Perspective on Turning Up the Immunological Heat on Cancer. *Front. Immunol.* **10**, 1654 (2019). doi: 10.3389/fimmu.2019.01654; pmid: 31379850
62. Y. H. Park *et al*., Chemotherapy induces dynamic immune responses in breast cancers that impact treatment outcome. *Nat. Commun.* **11**, 6175 (2020). doi: 10.1038/s41467-020-19933-0; pmid: 33268821
63. L. Zitvogel, L. Apetoh, F. Ghiringhelli, G. Kroemer, Immunological aspects of cancer chemotherapy. *Nat. Rev. Immunol.* **8**, 59–73 (2008). doi: 10.1038/nri2216; pmid: 18097448
64. I. D'Annessa, F. Cicconardi, D. Di Marino, Handling FMRP and its molecular partners: Structural insights into Fragile X Syndrome. *Prog. Biophys. Mol. Biol.* **141**, 3–14 (2019). doi: 10.1016/j.pbiomolbio.2018.07.001; pmid: 30905341
65. E. Pasciuto, C. Bagni, SnapShot: FMRP mRNA targets and diseases. *Cell* **158**, 1446–1446.e1 (2014). doi: 10.1016/j.cell.2014.08.035; pmid: 25215498
66. M. S. Taha *et al*., Novel FMRP interaction networks linked to cellular stress. *FEBS J.* **288**, 837–860 (2021). doi: 10.1111/febs.15443; pmid: 32525608
67. C. R. Hale *et al*., FMRP regulates mRNAs encoding distinct functions in the cell body and dendrites of CA1 pyramidal neurons. *eLife* **10**, e71892 (2021). doi: 10.7554/eLife.71892; pmid: 34939924
68. C. Schafroth *et al*., VE1 immunohistochemistry predicts *BRAF* V600E mutation status and clinical outcome in colorectal cancer. *Oncotarget* **6**, 41453–41463 (2015). doi: 10.18632/oncotarget.6162; pmid: 26496026
69. M. Wartenberg *et al*., Integrated Genomic and Immunophenotypic Classification of Pancreatic Cancer Reveals Three Distinct Subtypes with Prognostic/Predictive Significance. *Clin. Cancer Res.* **24**, 4444–4454 (2018). doi: 10.1158/1078-0432.CCR-17-3401; pmid: 29661773
70. J. Staaf *et al*., Whole-genome sequencing of triple-negative breast cancers in a population-based clinical study. *Nat. Med.* **25**, 1526–1533 (2019). doi: 10.1038/s41591-019-0582-4; pmid: 31570822
71. G. Sflomos, M. Shamsheddin, C. Brisken, An *ex vivo* model to study hormone action in the human breast. *J. Vis. Exp.* **95**, e52436 (2015). doi: 10.3791/52436; pmid: 25590412
72. P. Mali *et al*., RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013). doi: 10.1126/science.1232033; pmid: 23287722
73. Y. Guo *et al*., Metabolic reprogramming of terminally exhausted $CD8^+$ T cells by IL-10 enhances anti-tumor immunity. *Nat. Immunol.* **22**, 746–756 (2021). doi: 10.1038/s41590-021-00940-2; pmid: 34031618
74. G. Galliverti *et al*., Myeloid Cells Orchestrate Systemic Immunosuppression, Impairing the Efficacy of Immunotherapy against HPV^+ Cancers. *Cancer Immunol. Res.* **8**, 131–145 (2020). doi: 10.1158/2326-6066.CIR-19-0315; pmid: 31771984
75. C. Cianciaruso *et al*., Molecular Profiling and Functional Analysis of Macrophage-Derived Tumor Extracellular Vesicles. *Cell Rep.* **27**, 3062–3080.e11 (2019). doi: 10.1016/j.celrep.2019.05.008; pmid: 31167148
76. I. Keklikoglou *et al*., Chemotherapy elicits pro-metastatic extracellular vesicles in breast cancer models. *Nat. Cell Biol.* **21**, 190–202 (2019). doi: 10.1038/s41556-018-0256-3; pmid: 30598531
77. E. Elyada *et al*., Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. *Cancer Discov.* **9**, 1102–1123 (2019). doi: 10.1158/2159-8290.CD-19-0094; pmid: 31197017
78. S. Cheng *et al*., A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell* **184**, 792–809.e23 (2021). doi: 10.1016/j.cell.2021.01.010; pmid: 33545035
79. M. Foroutan *et al*., Single sample scoring of molecular phenotypes. *BMC Bioinformatics* **19**, 404 (2018). doi: 10.1186/s12859-018-2435-4; pmid: 30400809
80. M. Andreatta, A. J. Berenstein, S. J. Carmona, scGate: Marker-based purification of cell types from heterogeneous single-cell RNA-seq datasets. *Bioinformatics* **38**, 2642–2644 (2022). doi: 10.1093/bioinformatics/btac141; pmid: 35258562
81. D. A. Barbie *et al*., Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009). doi: 10.1038/nature08460; pmid: 19847166
82. S. R. Hingorani *et al*., Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* **7**, 469–483 (2005). doi: 10.1016/j.ccr.2005.04.023; pmid: 15894267
83. M. Germann *et al*., Neutrophils suppress tumor-infiltrating T cells in colon cancer via matrix metalloproteinase-mediated activation of TGFβ. *EMBO Mol. Med.* **12**, e10681 (2020). doi: 10.15252/emmm.201910681; pmid: 31793740
84. I. G. Maroulakou, M. Anver, L. Garrett, J. E. Green, Prostate and mammary adenocarcinoma in transgenic mice carrying a rat C3(1) simian virus 40 large tumor antigen fusion gene. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11236–11240 (1994). doi: 10.1073/pnas.91.23.11236; pmid: 7972041
85. Q. Zeng *et al*., Aberrant hyperexpression of the RNA binding protein FMRP in tumors mediates immune evasion, Zenodo (2022); https://doi.org/10.5281/zenodo.5159301.

ACKNOWLEDGMENTS

We wish to thank M. De Palma (EPFL), J. Schlessinger (Yale University), and G. Bollag (Opna Bio SA) for insightful discussions; M. De Palma (EPFL) for providing the OT1 mice and B16-OVA cells; G. Ciriello (University of Lausanne) for constructive comments and advice on the bioinformatic analyses; Y. Mohammadzadeh and B. Torchia (De Palma laboratory, EPFL) for helpful suggestions on the immune-related assays and exosome isolation, respectively; the patients, clinicians, and hospital staff members who contributed to the construction of the human tumor TMAs in Lund (TNBC, SCAN-B) and Bern (all other TMAs) as well as the staff at the central SCAN-B laboratory at Division of Oncology, Lund University, the Swedish National Quality Register for Breast Cancer (NKBC), Regional Cancer Center South, and the South Swedish Breast Cancer Group (SSBCG) for providing the SCAN-B TMA and associated transcriptome dataset, and in particular S. Lehn and P. Bolivar for liaison; L. Noti (University of Bern) for assisting in the analysis of human tumor TMAs; A. Ayyanan and C. Brisken (EPFL) for providing normal human breast tissue samples; A. Bowler (Radtke laboratory, EPFL) for providing mouse colon adenocarcinoma samples; L. Tang (EPFL) for support of Y.-Q.X.; M.-W. Peng, M. Gaveta, and E. Drori (EPFL) for technical support; all the members of the Flow Cytometry Core (especially V. Glutz for helping set up the FACS panels), the Bioimaging and Optics Core Facility, the Gene Expression Core Facility, the Histology Facility, and the Animal Care Facility (School of Life Sciences, EPFL) for respective technology support; and all the members of the Hanahan laboratory for discussions. **Funding:** The research was supported by grants from the Swiss National Science Foundation, by the Ludwig Institute for Cancer Research (New York), by a grant from the Biltema Foundation (Lund) administered by the ISREC Foundation (Lausanne), and by grants from the Cancera and Paulsson Foundations (Lund) and from Goran Grosskopf (Lund), all administered by the CHUV Foundation, Lausanne. **Author contributions:** Q.Z. contributed to the conception, experimental design, and implementation throughout; prepared figures; and edited the manuscript. S.S. performed bioinformatic analyses of RNA-seq and scRNA-seq datasets, developed FMRP activity signatures and their use in human association studies, prepared figures, and edited the manuscript. A.C. assessed and functionally validated adaptive and innate immune cell populations and their regulation, performed the macrophage reprogramming experiments, and prepared the figures and illustrations. N.F. contributed to bioinformatic analyses of RNA-seq and scRNA-seq datasets and the analyzing of TCR data. J.G. performed the mixing experiment shown in fig. S6. S.Y. contributed to the *Myc* KD experiment shown in fig. S3, I to K. P.L. queried the Encode ChIP-seq database for *Fmr1* as a target gene for MYC and MAX binding in cancer cell lines versus normal cell lines shown in fig. S3, D and E. L.C. analyzed human TMAs stained to detect FMRP expression shown in Fig. 1 and fig. S2. J.A.G. immunostained multiple human tumor TMAs and tissue sections for FMRP and participated in their analysis for Fig. 1. E.K. generated the PDAC TMA and contributed to its initial analysis. I.Z. supervised the analysis of multiple human tumor TMAs and tissue sections for FMRP expression shown in Fig. 1 and fig. S2. D.A. and F.G. contributed to the data shown in fig. S3, I to K; fig. S9H; and fig. S10, F and H. P.Z. participated in initial bioinformatic analysis of transcriptome datasets. Y.-Q.X. performed the CTL assay involving B6-OVA cancer cells and OT1 T cells. J.A.R.-C. and M.R. produced and provided the unpublished prostate cancer TMA analyzed in Fig. 1 and fig. S2. M.T. participated in FACS profiling of immune cell populations. K.H. contributed to the scRNA-seq. D.H. contributed to the conception, experimental design, project supervision, and the writing and preparation of the manuscript. **Competing interests:** Q.Z., S.S., D.A., and F.G. are currently employees of Opna Bio SA (Lausanne), which has licensed a patent application from EPFL describing FMRP as a cancer cell–intrinsic immunosuppressor and potential therapeutic target, in which Q.Z. and D.H. are coinventors. D.H. is a scientific founder of Opna Bio, and D.H., Q.Z., S.S., D.A., and F.G. have stock or stock options in the company. EPFL has filed a second patent application covering the prognostic FMRP cancer network signatures described herein, in which S.S. and D.H. are coinventors. No other authors have competing interests. **Data and materials availability:** Materials are available, subject to material transfer agreement (MTA) requests submitted to D.H. All data are available in the manuscript or the supplementary materials. Bulk RNA-seq and scRNA-seq, TCR sequencing, and whole-exome sequencing data in this study are deposited on Zenodo (*85*). **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl7207
Figs. S1 to S13
Tables S1 to S7
References (*86*, *87*)
MDAR Reproducibility Checklist

Submitted 4 August 2021; resubmitted 20 April 2022
Accepted 17 October 2022
10.1126/science.abl7207

Your Legacy to Science

AN ESTATE GIFT TO THE
AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE





Since 1848, our founding year, the American Association for the Advancement of Science (AAAS) has been deeply committed to advancing science, engineering and innovation around the world for the benefit of all people.

By making AAAS a beneficiary of your will, trust, retirement plan or life insurance policy, you become a member of our 1848 Society, joining Thomas Edison, Alexander Graham Bell and the many distinguished individuals whose vision led to the creation of AAAS and our world-renowned journal, *Science*, so many years ago.

Unlike many of its peers, *Science* is not for-profit. Your estate gift would provide long-term financial stability and durable annual income that will support operations and competitive innovation for years to come. **This support is vital.**

"As a teacher and instructor, I bear responsibility for the younger generations. If you have extra resources, concentrate them on organizations, like AAAS, that are doing work for all."

—Prof. Elisabeth Ervin-Blankenheim, 1848 Society member

If you intend to include AAAS in your estate plans, provide this information to your lawyer or financial adviser:

Legal Name: American Association for the Advancement of Science
Federal Tax ID Number: 53-0196568
Address: 1200 New York Avenue, NW, Washington, DC 20005

If you would like more information on making an estate gift to AAAS, cut out and return the form below or send an email to philanthropy@aaas.org. Additional details are also available online at **www.aaas.org/1848Society**.

1848 SOCIETY
AAAS

cut here

Yes, I would like more information about joining the AAAS 1848 Society.

PLEASE CONTACT ME AT:

Name: ______________________________

Address: ______________________________

City: ______________ State: ______ Zip code: ______________ Country: ______

Email: ______________________ Phone: ______________________

RETURN THIS FORM TO:
AAAS Office of Philanthropy and Strategic Partnerships • 1200 New York Avenue, NW • Washington, DC 20005 USA

RESEARCH ARTICLES

SOLAR CELLS

Initializing film homogeneity to retard phase segregation for stable perovskite solar cells

Yang Bai[1,2]†, Zijian Huang[3]†, Xiao Zhang[1]†, Jiuzhou Lu[1]†, Xiuxiu Niu[1]†, Ziwen He[4], Cheng Zhu[1], Mengqi Xiao[1], Qizhen Song[1], Xueyuan Wei[1], Chenyue Wang[1], Zhenhua Cui[1], Jing Dou[1], Yihua Chen[1], Fengtao Pei[1], Huachao Zai[3], Wei Wang[4], Tinglu Song[1], Pengfei An[5], Jing Zhang[5], Juncai Dong[5], Yiming Li[6], Jiangjian Shi[6], Haibo Jin[1], Pengwan Chen[7], Yuchao Sun[8], Yujing Li[1], Haining Chen[9], Zhongming Wei[10], Huanping Zhou[3], Qi Chen[1,2]*

The mixtures of cations and anions used in hybrid halide perovskites for high-performance solar cells often undergo element and phase segregation, which limits device lifetime. We adapted Schelling's model of segregation to study individual cation migration and found that the initial film inhomogeneity accelerates materials degradation. We fabricated perovskite films ($FA_{1-x}Cs_xPbI_3$; where FA is formamidinium) through the addition of selenophene, which led to homogeneous cation distribution that retarded cation aggregation during materials processing and device operation. The resultant devices achieved enhanced efficiency and retained >91% of their initial efficiency after 3190 hours at the maximum power point under 1 sun illumination. We also observe prolonged operational lifetime in devices with initially homogeneous $FACsPb(Br_{0.13}I_{0.87})_3$ absorbers.

With an endless compositional space that is almost continuously tailorable in the full region, halide perovskite materials present tunable electronic and optical properties (*1*–*5*). Although perovskite solar cells can have high power conversion efficiency (PCE), their limited device lifetime remains a substantial challenge to commercialization (*6*–*8*). The use of mixed cations and anions has led to a more appropriate tolerance factor of the crystal for improved structural stability and extra functionality to strengthen chemical stability in the resultant absorber thin films (*9*–*11*). However, mixed perovskite absorbers often undergo element and phase segregation that can decrease device efficiency and lifetime (*12*, *13*).

Most studies of phase segregation of mixed perovskites focus on film aging to understand cation and anion migration, the formation and development of nanoscale clusters (*14*–*16*), thermodynamic driving forces (*17*, *18*), and their impacts on film properties and device performance (*12*, *19*–*23*). Effective strategies such as relaxing residual strain and incorporating low-dimension perovskites can retard phase segregation by suppressing ion migration (*5*, *24*, *25*). However, relevant metrics are needed to investigate the atomistic aggregation of individual ions to correlate their collective behavior that led to film degradation.

We adapted a physical analog of Schelling's model, which was used to illustrate agent (such as isolated particles) segregation caused by even low levels of individual preference in the context of social economics (*26*) and physics (*27*). It bridges the microscale analysis of cation aggregation in perovskites and macroscale observations of their phase separation and film degradation. In the simulation, the initial film homogeneity affected cation aggregation duration and thus device lifetime. On the basis of these findings, we grew homogeneous α-phase $FA_{1-x}Cs_xPbI_3$ (FA, formamidinium) that resulted in a solar cell with enhanced efficiency and prolonged operational lifetime. Furthermore, for the wide-bandgap $FACsPb(Br_{0.13}I_{0.87})_3$, the corresponding device retained >80% of its initial PCE after 340 hours at maximum power point (MPP) tracking. It indicates that this model is applicable to various perovskite formulations such as mixed halide perovskites, which demonstrates the generality of the Schelling model for halide perovskites.

Cation aggregation and phase transition

The α-phase $FA_xCs_{1-x}PbI_3$ (FACs) perovskites have an optimal bandgap and high thermal stability from the resultant crystal structure, with an ideal Goldschmidt tolerance factor (*11*, *21*, *28*–*31*). However, reported devices based on FACs perovskite absorbers have had limited lifetimes, resulting from material degradation that is largely the result of phase separation (*32*). In Fig. 1, A, C, and D, we show the two-dimensional (2D) photoluminescence (PL) mapping by wavelength, intensity, and full width at half-maximum (FWHM), respectively, of the $FA_{0.9}Cs_{0.1}PbI_3$ perovskite film. The emission spectra display peak variability between the selected regions. Moreover, although the emission of region 1 was concentrated at 795 nm (attributed to $FA_{0.9}Cs_{0.1}PbI_3$) (*12*), both regions 2 and 3 exhibited splitting of the emission peak (Fig. 1B). This feature was indicative of phase segregation within the perovskite film.

We investigated the segregation evolution of films aged under continuous light illumination for 20 min at 80°C at the same location in ambient atmosphere. As shown in the mapping spectra (Fig. 1E), the phase segregation of the film intensified with increasing areas of both the yellow domains (which represent the longer-wavelength, FA-rich composition) and red domains (which represent the shorter-wavelength, Cs-rich composition), which indicated that the local Cs/FA ratio deviated from that observed in the pristine films. Region 1 showed peak splitting at 799 and 789 nm, with increased peak intensity at longer wavelengths. By contrast, for regions 2 and 3, the peak intensity increase was observed at shorter wavelengths (783 and 781 nm, respectively). Moreover, the FWHM increased notably (Fig. 1H), which indicates the broadening of the PL emission peak after aging. After comparison with the pure $FAPbI_3$ film, for which no obvious change on the PL mapping was observed after aging (fig. S1), we hypothesized that Cs^+ cations in region 1 migrated to regions 2 and 3 in the mixed-cations composition of the perovskite with a single halide. This hypothesis was confirmed through time-of-flight secondary-ion mass spectrometry (TOF-SIMS) mapping (fig. S2).

Upon extending the aging duration to 1000 hours (supplementary materials), we observed crystal phase transformation in the film. In addition to the cubic perovskite peak (14.0°), two extra diffraction peaks at 9.7° and 11.6° were dominant in the x-ray diffraction (XRD) patterns (figs. S3 and S4), which indicates

[1]Beijing Key Laboratory of Construction Tailorable Advanced Functional Materials and Green Applications, MIIT Key Laboratory for Low-dimensional Quantum Structure and Devices, Experimental Center of Advanced Materials, School of Materials Science and Engineering, Beijing Institute of Technology, Beijing 100081, P. R. China. [2]Advanced Research Institute of Multidisciplinary Science, Beijing Institute of Technology, Beijing 100081, P. R. China. [3]Department of Materials Science and Engineering, College of Engineering, Peking University, Beijing 100871, P. R. China. [4]Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, P. R. China. [5]Beijing Synchrotron Radiation Facility, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, P. R. China. [6]Key Laboratory for Renewable Energy, Beijing Key Laboratory for New Energy Materials and Devices, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, P. R. China. [7]State Key Laboratory of Explosion Science and Technology, Beijing Institute of Technology, Beijing 100081, P. R. China. [8]Auner Technology Co., Ltd., Beijing 100084, P. R. China. [9]School of Materials Science and Engineering, Beihang University, Beijing 100191, P. R. China. [10]State Key Laboratory of Superlattices and Microstructures, Institute of Semiconductors, Chinese Academy of Sciences and Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing 100083, P. R. China.
***Corresponding author. Email: qic@bit.edu.cn**
†These authors contributed equally to this work.

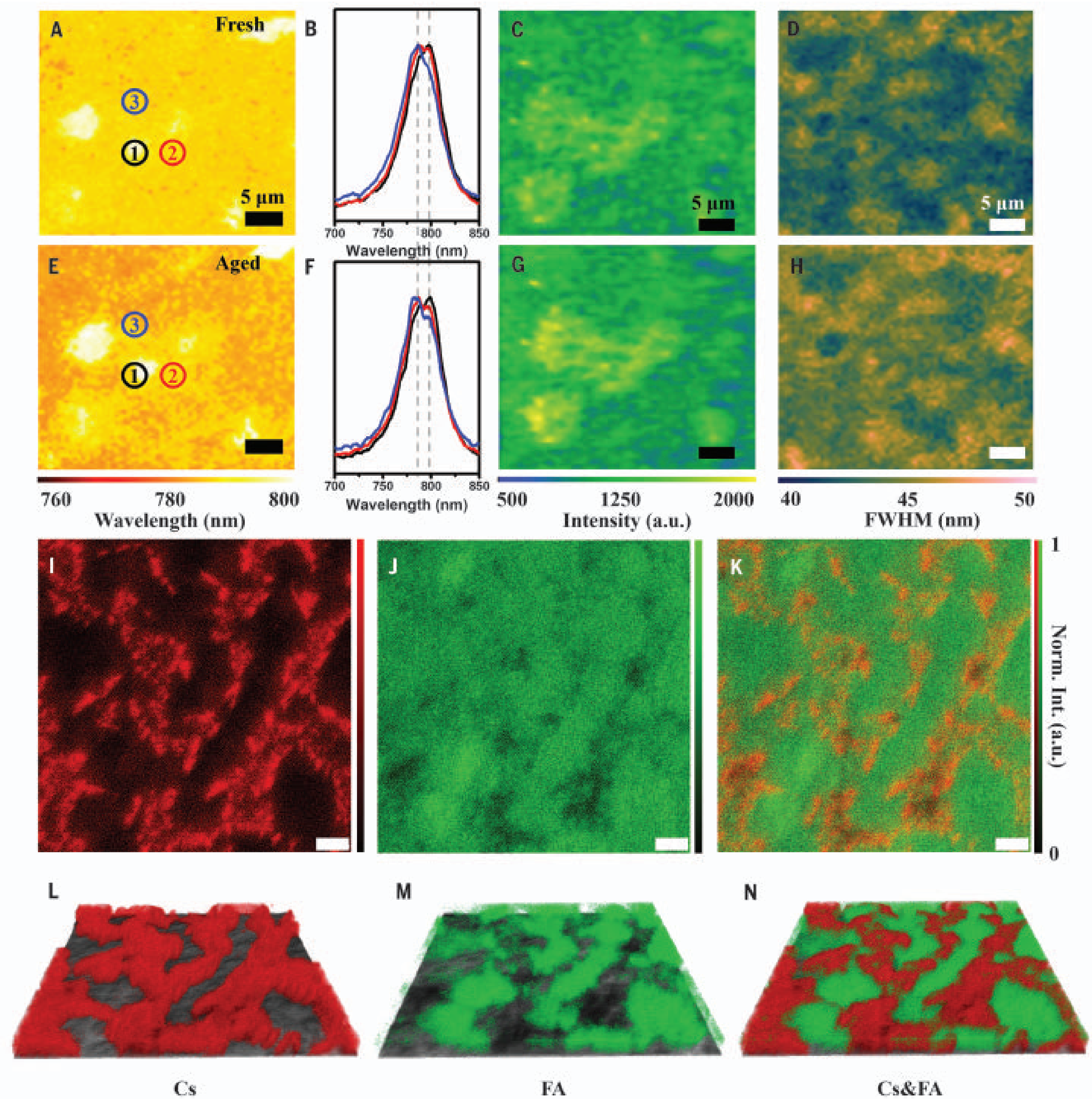


Fig. 1. Degradation mechanisms for binary FACs perovskites. (**A** to **H**) 2D PL mapping spectra (wavelength) of perovskite film before (A) and after (E) degradation. [(B) and (F)] PL spectra of selected regions from (A) and (E), respectively. [(C), (D), (G), and (H)] The corresponding 2D PL intensity and FWHM mapping spectra of (A) and (E). a.u., arbitrary units. (**I** to **K**) The (I) Cs and (J) FA intensity mapping for the degraded film measured with TOF-SIMS and (K) the overlay mapping spectra. Scale bar, 10 µm. (**L** to **N**) 3D topographic of Cs and FA for the degraded films measured by TOF-SIMS.

the formation of nonphotoactive δ-$CsPbI_3$ and δ-$FAPbI_3$ phases (*18*, *33*, *34*). To correlate the composition of each region to its crystal structure, we conducted 2D TOF-SIMS mapping on the fully degraded film. The Cs distribution (Fig. 1I) showed domains at the scale of several to several tens of micrometers, which occurred in parallel to FA segregation (Fig. 1J). The separated Cs/FA domains were also spatially complementary as expected (Fig. 1K), given the binary cation nature of the perovskite system. Moreover, depth-dependent TOF-SIMS measurement (Fig. 1, L to N) showed that vertical direction distribution was relatively uniform, so cations at different depths of the bulk aggregated in a manner similar as that at the surface (fig. S5).

Cation segregation dynamics by the Schelling model

On the basis of the XRD, PL mapping, and TOF-SIMS results, we argue that the degradation of the binary system follows the route

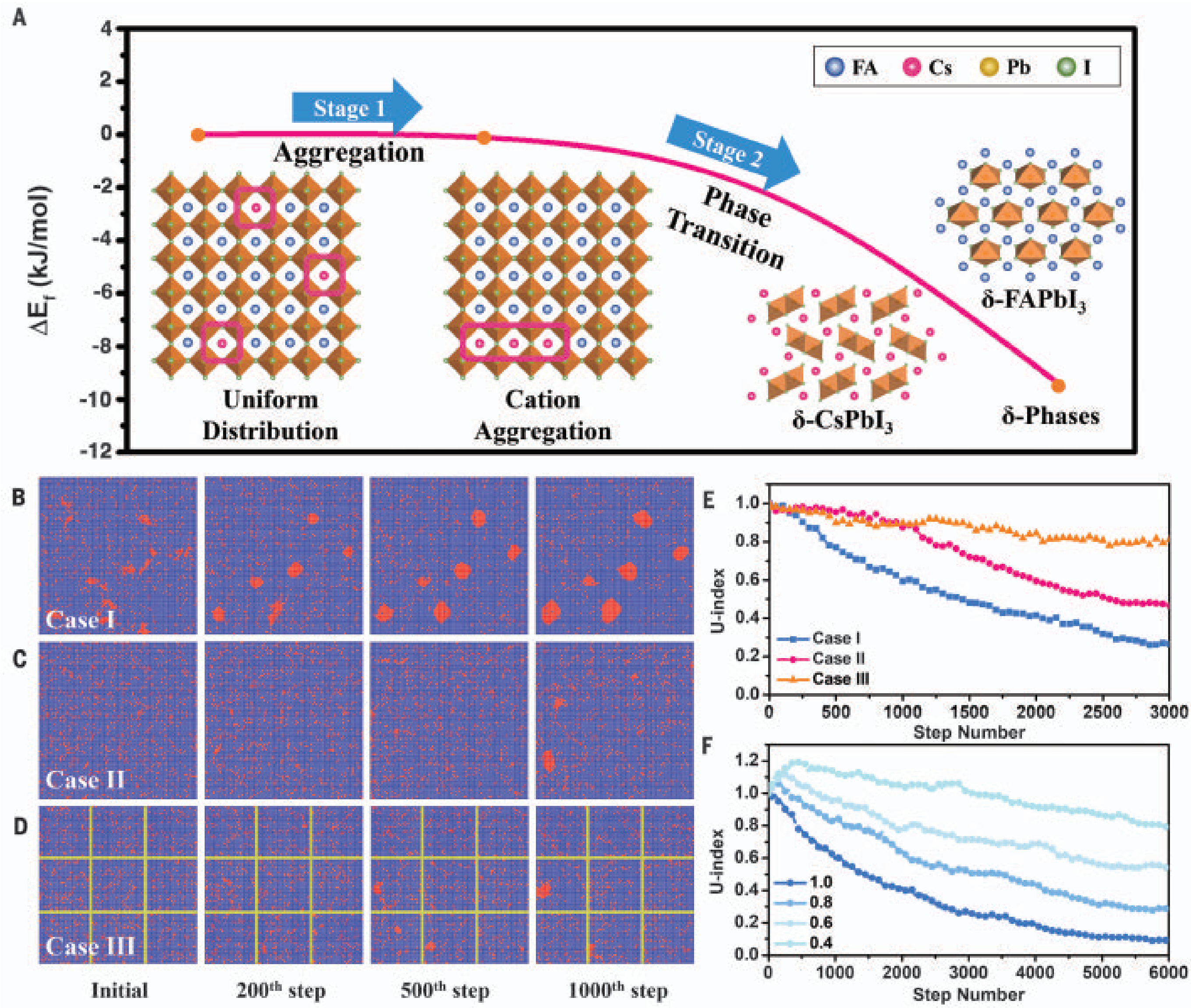


Fig. 2. Schelling model simulations of cation segregation dynamics. (**A**) Energy change of the degradation processes from a uniform distribution to phase segregation to δ-phases, by means of DFT. (**B** to **D**) Simulations of segregation process on the basis of the Schelling model. (B) An initially segregated distribution (case I, ISP = 27%). (C) An initially less segregated distribution (case II, ISP = 2%). (D) A 1D and 3D (or 2D and 3D) fused system (case III, ISP = 2%). (**E** and **F**) U index evolution of red particles (indicates Cs) derived from (E) cases I to III and (F) cases with different moving tendencies.

from cation aggregation to phase transition (Fig. 2A). Local aggregation of cations initiates phase transition that leads to the complete process of phase segregation. The thermodynamic driving force for the phase segregation was investigated by means of density functional theory (DFT) calculations (fig. S6). DFT results identified that the $FA_{0.89}Cs_{0.11}PbI_3$ perovskite exhibited an energy change of only −0.133 kJ mol^{-1} from the homogeneous distribution to cation aggregation (Fig. 2A, stage 1), which is possibly due to the reduction in interfaces between the FA and Cs domains. However, in the subsequent process, the segregated domains experienced phase transition to form δ-$FAPbI_3$ and δ-$CsPbI_3$ phases (Fig. 2A, stage 2), in which the energy change of −9.427 kJ mol^{-1} acted as the energy sink corresponding for the irreversible segregation. Our observations showed that cation aggregation occurred before phase transition, which provides an opportunity to intervene before δ-phase formation occurs (*20*, *21*).

The Schelling model of segregation can be applied to isolated particles (such as molecules or ions) on the basis of a few simple physical assumptions (see supplementary text S1) (*27*). We adapted the Schelling model to study the cation dynamics in perovskite films by means of a Monto Carlo algorithm (fig. S7). Given that cations at different depths of the film aggregated in the same manner as at the surface (Fig. 1, I to N), we simplified the simulation to a 2D case (*35*, *36*) with a 100 by 100 lattice to accommodate binary cations of Cs (Fig. 1, red particles) and FA (Fig. 1, blue particles) that could randomly swap positions with neighbors.

In the simulation for any step, each single cation has a probability to exchange its position with neighbors. We defined this probability as moving tendency that is the product of the thermodynamic parameter of the phase transition (η_T) and kinetic parameter of cation aggregation (η_K). DFT results revealed the phase formation energy to identify η_T in the Schelling model. Particularly, η_T values were assigned to be 1 for all spontaneous processes with a negative value for energy change. The kinetic parameter η_K was determined by the

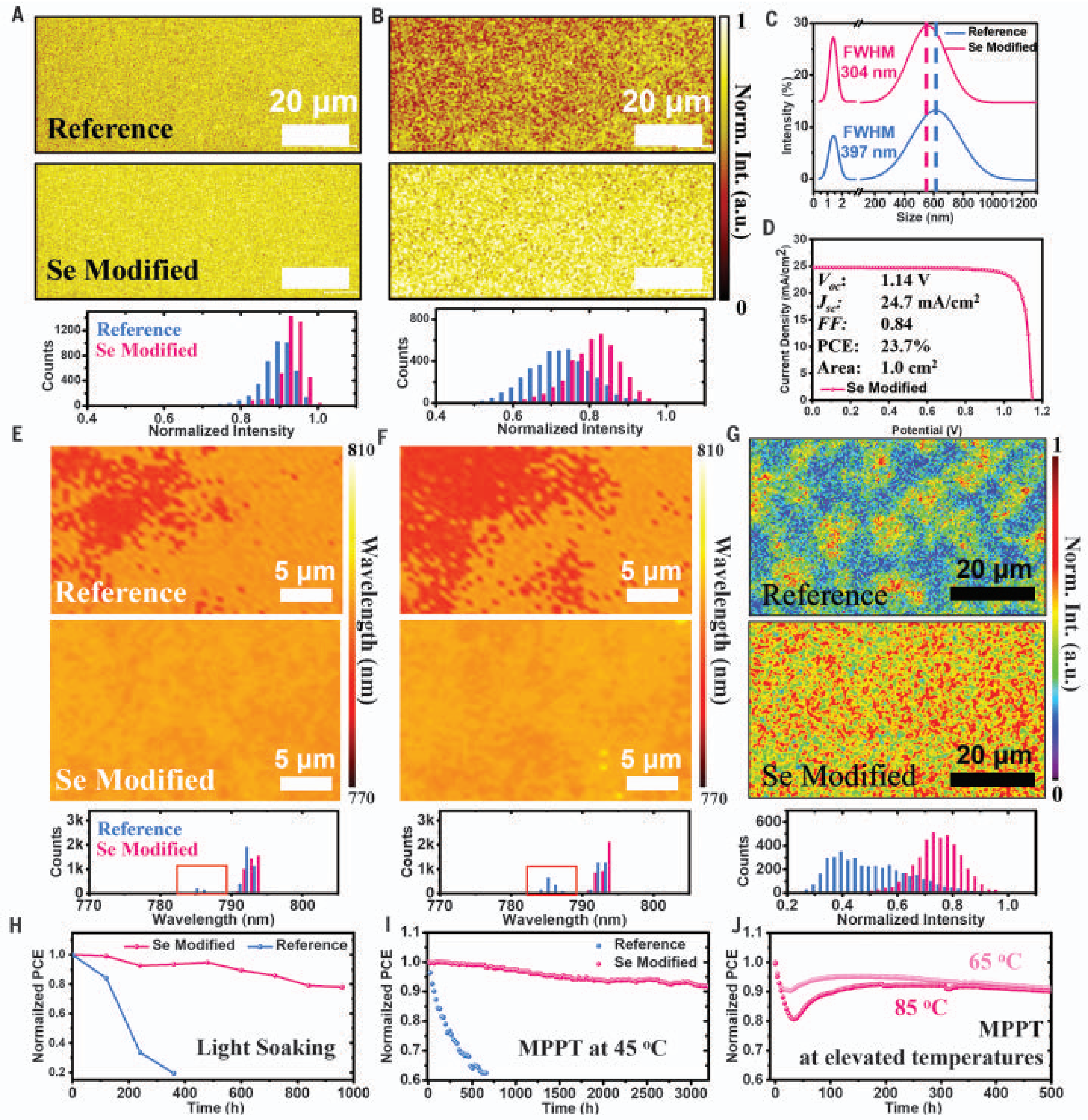


Fig. 3. Film homogeneity of mixed cation perovskite and the impacts. (**A** and **B**) 2D TOF-SIMS mapping of Cs^+ for (A) the first step of PbI_2/CsI film fabricated by reference and selenophene-modified precursor solutions and (B) perovskite films fabricated by reference and modified precursor solutions, as well as the corresponding intensity distribution. (**C**) Colloid size distribution in the perovskite precursor solutions obtained from DLS. (**D**) *J-V* curves of the champion devices from homogeneous film with an active area of 1 cm^2. (**E** and **F**) In situ 2D PL mapping spectra evolution of perovskite films (E) before and (F) after aging under 80°C thermal treatment in an ambient environment with a relative humidity of 40% and the corresponding statistics diagrams. (**G**) 2D TOF-SIMS mappings of perovskite films (Cs) after continuous thermal treatment in an N_2-filled glovebox at 85°C for 144 hours with the corresponding statistics diagrams. (**H**) Time-dependent PCE measurements under continuous light illumination (LED source with 100 mW/cm^2) held at the open circuit in an N_2 atmosphere. (**I** and **J**) MPP tracking of devices under 100 mW/cm^2 LED source in an N_2-filled glovebox without encapsulation at (I) 45° ± 5°C and (J) elevated temperatures of 65° ± 5°C and 85° ± 5°C. MPPT, MPP tracking.

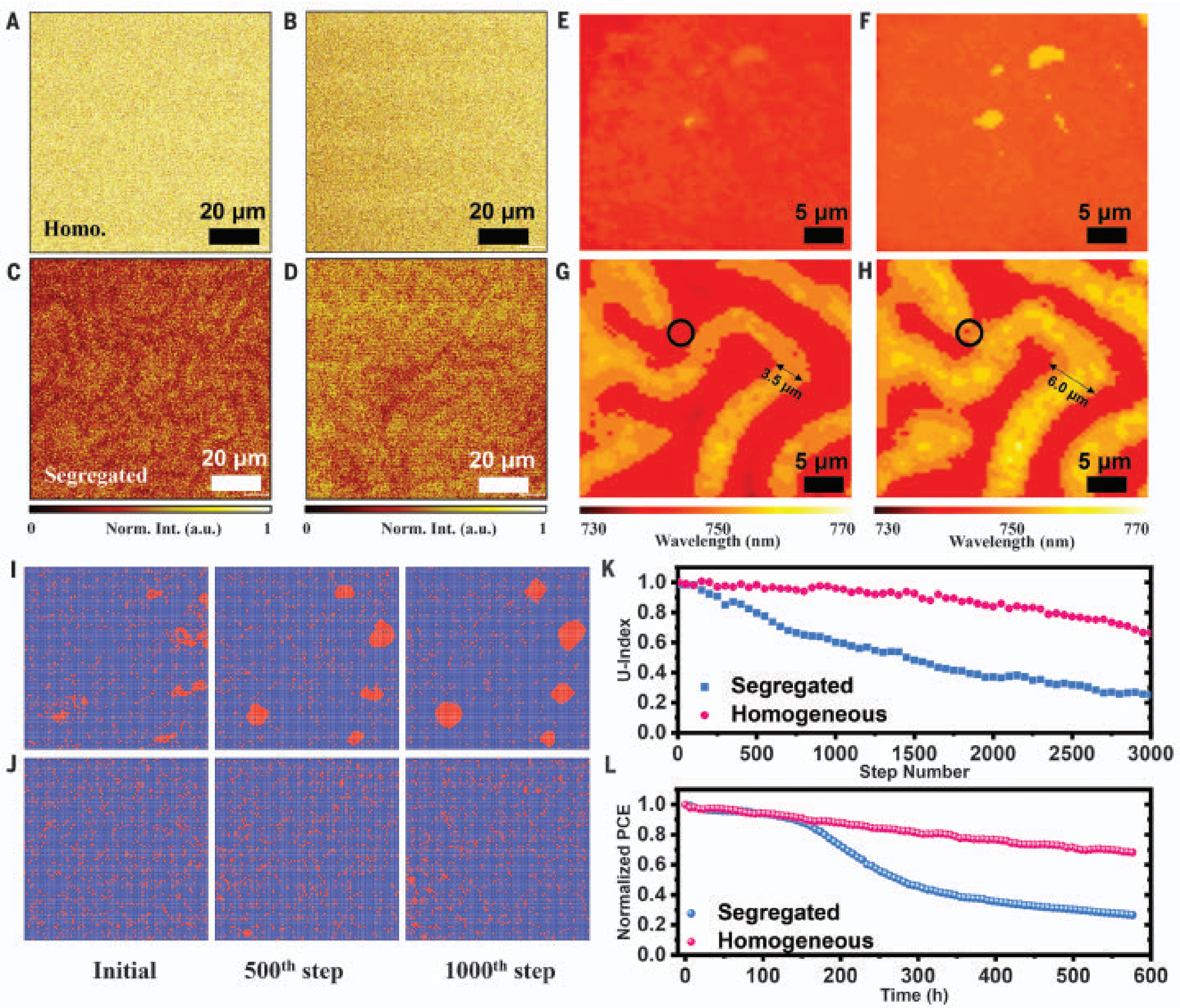


Fig. 4. Film homogeneity of mixed halide perovskite and the impacts. (**A** to **D**) 2D TOF-SIMS mapping of [(A) and (C)] I^- in the mixed halide perovskite films (A) without and (C) with initial segregation and [(B) and (D)] perovskite films aged under continuous light illumination at 45° ± 5°C in an N_2-filled glovebox for 24 hours (B) without and (D) with severe initial iodide segregation. (**E** to **H**) In situ 2D PL mapping spectra (wavelength) evolution of perovskite films [(E) and (F)] without and [(G) and (H)] with severe iodide segregation, under light illumination for 20 min in an ambient environment with relative humidity of 60%. (**I** and **J**) Simulation of segregation process of perovskite films in (I) an initially segregated distribution (ISP = 28%) and (J) an initially less segregated distribution (ISP = 3%). (**K**) U index evolution of red particles (Br) derived from simulation cases. (**L**) MPP tracking under 100 mW/cm^2 LED illumination in an N_2-filled glovebox without encapsulation.

ion migration activation energy, depending on the cation properties and their local chemical environment. In a particular scenario for simulation, we simply normalized the migration activation energy for different cations from DFT or experimental results to obtain η_K ranging from 0 to 1. To provide a quantitative measure of the film homogeneity, we defined unsegregated index (U index) as the ratio of presently unsegregated Cs ions to the initially unsegregated ones. Regarding any case of interest, the heavier segregation can be reflected on the smaller U index (Fig. 2, B and E).

The initial state of cation aggregation affected the phase segregation trajectories. We investigated phase segregation with different initial segregation percentages (ISPs), which stands for the ratio of initially segregated Cs to the total amount of Cs. The values of other core parameters are summarized in tables S1 and S2. The cation aggregation with an arbitrary ISP of 27% was simulated for clear demonstration (Fig. 2B, case I). After 200 steps, the U index decreased from 1.00 to 0.94. After 500 and 1000 steps, the U index dropped to 0.77 and 0.59, respectively, and heavier aggregation and enlarged domains were observed (movie S1).

We compared the initially more homogeneous film (case II, ISP of 2%) with the inhomogeneous one (case I) through simulation. After 200 steps, the U index for case II only dropped to 0.98 (Fig. 2C, middle left), in which

the aggregation is negligible as compared with the initial state. After 500 and 1000 steps, the U index remained at 0.95 and 0.88, respectively (movie S2). With linear regression, we obtained the U index decay rate of 0.012 per 100 steps for case II, whereas that for case I is 0.041 per 100 steps. The initial homogeneous film clearly exhibited slower segregation (Fig. 2, B and C, and fig. S8), which suggests a longer lifetime in the corresponding devices.

Increasing the activation energy of ion migration is proven to prolong device lifetime (*10*), which can be reflected as a lower η_K in our model. Simulations for the group of films with the same initial homogeneity (ISP of 27%) but decreased η_K (1, 0.8, 0.6, and 0.4) showed reduced decay rates of 0.041, 0.022, 0.017, and 0.007 per 100 steps in their linear decay regions, respectively (Fig. 2F, fig. S9, and table S2). The initial increase of U index could be observed in some cases, in which a large number of small segregation domains were dissolved to form larger ones (supplementary materials). It reveals the determinant impact of the initial state on the segregation dynamics, which were in good agreement with the experimental data reported elsewhere (*23*).

We also simulated the aggregation dynamics of ternary cation perovskites (fig. S10) and 1D and 3D fused perovskites (Fig. 2, D and E, and fig. S11). For the fused perovskites (case III), the immobile large cations created barriers between each sublattice and retarded cation migration. According to previous observations (*25*, *37*), we reasonably assumed the η_K of large cations (yellow) to be zero. These boundaries retarded phase segregation (movie S3), and the decay rate further declined to 0.006 per 100 steps, which was just half of that of case II and in agreement with reported experimental data (*37*, *38*). The stability enhancement in 1D and 3D perovskites is generally attributed to the prevention of moisture by large hydrophobic cations, but our findings suggest that retarded phase segregation is caused in part by partially immobilized cations.

Initializing film homogeneity and device performance

We modulated the Cs aggregation in FACs perovskite thin films with a two-step deposition process to introduce CsI when depositing the PbI_2 film in the first step (fig. S12) (*39*). The Coulomb interaction between the Cs^+ cation and the PbI_4^{2-}-based complex led to the formation of CsI–PbI_2 colloids in solution (fig. S13) and subsequent Cs-rich aggregates in the precursor film after thermal annealing at 70°C (*40*). The top-view mapping of Cs^+ on the PbI_2/CsI films that were characterized by TOF-SIMS (Fig. 3A) revealed the inhomogeneous distribution of Cs-rich domains on the micrometer scale. After the organic components were added, the CsI/PbI_2 was converted to perovskite films when annealed at 150°C. The inhomogeneous Cs^+ distribution was still observed with TOF-SIMS mapping (Fig. 3B) and the PL measurements (fig. S14).

For less inhomogeneity, we modulated the Pb-solvent interactions by introducing a donor molecule to reduce the CsI–PbI_2 colloid size (*41*). Common Lewis bases (such as dimethyl sulfoxide) lead to strong coordination with Pb and form crystalline intermediates (*42*). To achieve a moderate coordination with Pb and disturb those crystalline intermediates, two weak donors, thiophene and selenophene, were separately added to the PbI_2/CsI precursor solutions to grow films. The 2D TOF-SIMS mapping of both films revealed a more homogeneous distribution of Cs than that of the reference sample (Fig. 3A and fig. S15 for selenophene and thiophene, respectively), although the selenophene was more homogeneous.

The intensity distribution analysis of CsI–PbI_2 film (Fig. 3A) showed a reduced FWHM for the selenophene-modified film (0.04 versus 0.08 for the reference), which indicates the smaller intensity difference of Cs at each scanning point (~0.1 by 0.1 μm^2). In addition, the x-ray photoelectron spectroscopy (XPS) mapping (fig. S16) showed a narrow intensity variation in the modified film, which also demonstrated an enhanced film homogeneity. No phase structure difference was detected from the XRD patterns (fig. S17A). As expected, the resultant perovskite film (selenophene modified) was more uniform than the reference (Fig. 3B), with the FWHM of the perovskite peak from XRD patterns at 14.0° decreasing from 0.19 to 0.14 (fig. S17B). No obvious phase segregation was detected in the PL mapping measurement for the modified film (fig. S14).

We studied the interaction between selenophene and Pb in the precursor solution with several methods. Carbon nuclear magnetic resonance (^{13}C-NMR) revealed that the chemical shift of the α carbon in selenophene at 131.95 parts per million (ppm) moved downfield to 132.05 ppm, which is indicative of chemical bonding between selenium (Se) and PbI_2 (fig. S19). Extended x-ray absorption fine structure (EXAFS) spectra of the Pb L_3-edge (fig. S20 and table S3) showed that the coordination numbers of Pb in Pb-I before and after the introduction of selenophene were 4.2 ± 0.6 and 3.6 ± 0.5, respectively. The EXAFS results are not conclusive because of its limited sensitivity but are in line with the ^{13}C-NMR results, which is indicative of an interaction between Pb and Se in the precursor (*43*). Dynamic light scattering (DLS) showed that the mean colloidal size decreased from 622 nm in the reference to 552 nm (Fig. 3C). We argue that selenophene interacted with Pb to disturb the formation of large colloid clusters in the precursor. As compared with large clusters, smaller clusters effectively diluted the aggregation of Cs, which mitigated the heterogeneity in the resultant film.

A series of devices, with the configuration of indium-doped tin oxide (ITO)/SnO_2/$Cs_{0.1}FA_{0.9}PbI_3$/2,2′,7,7′-tetrakis[N,N-di(4-methoxyphenyl)amino]-9,9′-spirobifluorene (Spiro-OMeTAD)/Au, were fabricated by tuning the initial film homogeneity (fig. S22), and the current-voltage (*J-V*) curves of champion devices are shown in fig. S23 with the key parameters summarized in table S4. On the basis of the enhanced initial film homogeneity with selenophene, the device performance improved from 20.9 to 23.4% (0.08 cm^2, average PCE from 20.1 to 22.6%). We observed improved photoelectronic properties in the resultant device with suppressed carrier recombination (fig. S24), lower trap density (fig. S25), enhanced mobility (fig. S25), and better charge extraction (fig. S26). With further optimization, the device with an active area of 1.00 cm^2 showed a certified PCE of 23.7% (reverse scan), with an open-circuit voltage (V_{oc}) of 1.14 V, a short-circuit current density (J_{sc}) of 24.7 mA per cm^2, and a fill factor (*FF*) of 0.84 (Fig. 3D). The device was also subjected to continuous illumination by holding at 1.03 V and achieved a stabilized efficiency of 22.4% after 300 s in ambient atmosphere without encapsulation (fig. S27).

The absorber films with improved initial homogeneity (with selenophene) exhibited enhanced stability as compared with the reference film. We conducted in situ PL mapping measurements to the films during aging under thermal treatment. After 20 min of thermal aging in an ambient atmosphere at 80°C, the reference sample exhibited two emissive regions, of which the ones at ~785 nm grew larger (Fig. 3, E and F). This change was most likely caused by the formation of Cs-rich and Cs-poor domains, in which the local Cs-FA stoichiometry deviated from that in the pristine films. By contrast, the perovskite film with enhanced initial homogeneity exhibited a constant and uniform emission at 794 nm with retarded Cs aggregation. These results were confirmed by means of TOF-SIMS mapping (Fig. 3G) of films after aging for 144 hours. Scanning electron microscopy (SEM) images also revealed that the initially segregated film underwent substantial morphology changes after the thermal treatment, whereas no obvious changes were observed for the selenophene-treated film (fig. S28).

The homogeneous films also exhibited enhanced light stability. After aging under continuous illumination for 144 hours in an N_2-filled glovebox, notable morphology changes occurred in the reference at the grain boundaries as shown in SEM images (fig. S29, A and B), and the δ-$FAPbI_3$ phase was identified by means of XRD (fig. S29, C and D). By contrast, no

obvious morphology change occurred on the homogeneous films, and we observed a retarded formation of δ-$FAPbI_3$. The 2D PL mapping measurements show that the PL peak of reference film exhibited blue shift from 790 to 785 nm with the emergence of nonphotoactive regions, possibly owing to the formation of δ-$FAPbI_3$. However, the film with improved initial homogeneity showed only slight changes of the peak position and distribution, which illustrates the enhanced light stability (fig. S30).

The DFT simulations suggested that the remaining selenophene can benefit the suppressing of Cs migration with increased activation energy (fig. S31). XPS data showed that after being stored in the ambient environment for 12 hours, the perovskite films still exhibited the symmetric peaks of Pb (II) at 138.5 and 143.4 eV that are attributed to Pb $4f_{7/2}$ and Pb $4f_{5/2}$, respectively. It indicates that the film surface was protected from oxidation, which may have originated from the coordination of residual selenophene (fig. S32). However, the remaining amount is small according to the TOF-SIMS measurement (fig. S33), so the enhanced film stability could be mainly attributed to the increased initial film homogeneity.

Selenophene-treated perovskite devices were aged under various conditions according to the modified International Summit on Organic Photovoltaic Stability (ISOS) stability protocols (*44*) and presented enhanced stability than that of reference devices. The selenophene-treated perovskite device showed good stability without PCE degradation more than 2000 hours after being stored in the dark in an N_2 atmosphere (fig. S34). The devices that were aged at 85°C in N_2 in the dark (fig. S35) retained ~80% of their initial PCE after 1750 hours. When aged over 1-sun continuous illumination [light-emitting diode (LED) source, 100 mW/cm^2] at open circuit, the devices could retain ~80% of their initial PCE after 960 hours (Fig. 3H and fig. S36). For the MPP test at 45° ± 5°C in an N_2 atmosphere, the device retained 91% of its highest PCE after 3190 hours. In addition, the MPP test was performed at elevated temperatures in the N_2 atmosphere, and the devices retained 91 and 89% of their highest PCE after 500 hours at 65° ± 5°C and 85° ± 5°C, respectively (Fig. 3J and fig. S37).

For mixed-halide perovskites, the main device issue has been the poor light stability. We fabricated $FACsPb(Br_{0.13}I_{0.87})_3$ of a bandgap of 1.66 eV with different initial halide segregation. We compared 2D TOF-SIMS mapping on the absorber films, which were subjected to continuous light illumination for 24 hours at 45° ± 5°C in an N_2-filled glovebox. The film without severe initial segregation that was fabricated under the protocol that is reported elsewhere (*45*) (supplementary materials, materials and methods) retained a homogeneous I^- distribution (Fig. 4, A and B). However, for the film with heavier initial segregation, the larger I^--deficient regions (~5 μm) were identified on the resultant film (Fig. 4, C and D).

In addition, we conducted in situ PL mapping as films were being aged under light illumination. Similar to cation segregation, the film with initially homogeneous halide distribution showed better stability under continuous illumination (Fig. 4, E and F). By contrast, the film with heavier initial halide segregation showed a marked change on the PL mapping spectra (Fig. 4, G and H, and fig. S38). We observed that the width of longer-wavelength regions (Fig. 4, G and H, yellow) was increased from ~3 to 6 μm. Moreover, the emission peak was clearly redshifted from 739 to 751 nm (Fig. 4, G and H, circled region), which indicates the fast halide segregation. The Schelling model simulation results are in good agreement with the experimental observation. As shown in Fig. 4, I to K, the segregated case (ISP = 28%), presented a decay rate of 0.040 per 100 steps, which is more than four times higher than that of the initially homogeneous film of 0.008 per 100 steps (ISP = 3%).

We further fabricated devices based on these two absorbers, which showed PCEs of 18.78% (homogeneous films) and 17.93% (reference), respectively, as shown in fig. S39. The mixed halide device with initially homogeneous halide distribution achieved an impressive operational stability during the MPP test at 45° ± 5°C in N_2, which retained 80% of its initial PCE after 340 hours (168 hours for the reference device). To date, this stability result is among the best device operational lifetimes (T_{80}) based on mixed halide perovskites (*46*). In addition, at an elevated temperature of 85° ± 5°C, the mixed halide perovskite device with improved homogeneity achieved a T_{80} lifetime of 80 hours (fig. S40).

We have demonstrated that the Schelling model is a powerful tool to bridge theoretical analysis of perovskites at the atomic scale and macroscale observations of their phase separation and film degradation. From simulation and experiment results, we found that the initial film homogeneity, in terms of elemental distribution, has shown substantial influence on film and device stability. Benefiting from homogeneous films by tailoring precursor chemistry with selenophene, we developed high-performance devices based on mixed perovskites that achieved a long-term stability during the MPP test even at elevated temperatures.

REFERENCES AND NOTES

1. L. Gu *et al*., *Nature* **581**, 278–282 (2020).
2. K. Lin *et al*., *Nature* **562**, 245–248 (2018).
3. Z. Xiao *et al*., *Nat. Mater.* **14**, 193–198 (2015).
4. M. Jeong *et al*., *Science* **369**, 1615–1620 (2020).
5. G. Kim *et al*., *Science* **370**, 108–112 (2020).
6. S. Bai *et al*., *Nature* **571**, 245–250 (2019).
7. E. H. Jung *et al*., *Nature* **567**, 511–515 (2019).
8. H. Min *et al*., *Nature* **598**, 444–450 (2021).
9. M. Abdi-Jalebi *et al*., *Nature* **555**, 497–501 (2018).
10. N. X. Li *et al*., *Nat. Energy* **4**, 408–415 (2019).
11. S. H. Turren-Cruz, A. Hagfeldt, M. Saliba, *Science* **362**, 449–453 (2018).
12. N. X. Li *et al*., *Joule* **4**, 1743–1758 (2020).
13. J. P. Correa-Baena *et al*., *Science* **363**, 627–631 (2019).
14. T. A. S. Doherty *et al*., *Nature* **580**, 360–366 (2020).
15. S. Macpherson *et al*., *Nature* **607**, 294–300 (2022).
16. K. Frohna *et al*., *Nat. Nanotechnol.* **17**, 190–196 (2022).
17. J. Barrier *et al*., *Energy Environ. Sci.* **14**, 6394–6405 (2021).
18. L. T. Schelhas *et al*., *Energy Environ. Sci.* **12**, 1341–1348 (2019).
19. D. B. Straus, S. Guo, A. M. M. Abeykoon, R. J. Cava, *Adv. Mater.* **32**, e2001069 (2020).
20. S. Kim, T. Eom, Y. S. Ha, K. H. Hong, H. Kim, *Chem. Mater.* **32**, 4265–4272 (2020).
21. C. Y. Yi *et al*., *Energy Environ. Sci.* **9**, 656–662 (2016).
22. C. A. R. Perini, T. A. S. Doherty, S. D. Stranks, J.-P. Correa-Baena, R. L. Z. Hoye, *Joule* **5**, 1024–1030 (2021).
23. L. E. Mundt *et al*., *ACS Energy Lett.* **7**, 471–480 (2022).
24. Y. H. Lin *et al*., *Science* **369**, 96–102 (2020).
25. Y. Lin *et al*., *ACS Energy Lett.* **2**, 1571–1572 (2017).
26. T. C. Schelling, *J. Math. Sociol.* **1**, 143–186 (1971).
27. D. Vinkovic, A. Kirman, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19261–19265 (2006).
28. M. Saliba *et al*., *Science* **354**, 206–209 (2016).
29. Z. Li *et al*., *Chem. Mater.* **28**, 284–292 (2016).
30. M. M. Hao *et al*., *Nat. Energy* **5**, 79–88 (2020).
31. S. Kawachi *et al*., *J. Phys. Chem. Lett.* **10**, 6967–6972 (2019).
32. K. Ho, M. Wei, E. H. Sargent, G. C. Walker, *ACS Energy Lett.* **6**, 934–940 (2021).
33. Y. Liu *et al*., *Angew. Chem. Int. Ed.* **59**, 15688–15694 (2020).
34. S. S. Xiang *et al*., *ACS Energy Lett.* **3**, 1824–1831 (2018).
35. H. Pan *et al*., *Science* **374**, 100–104 (2021).
36. W. Mao *et al*., *Nat. Mater.* **20**, 55–61 (2021).
37. M. Parashar, R. Singh, K. Yoo, J.-J. Lee, *ACS Appl. Energy Mater.* **4**, 2751–2760 (2021).
38. Z. Wang *et al*., *Nat. Energy* **2**, 17135 (2017).
39. J. H. Im, I. H. Jang, N. Pellet, M. Grätzel, N. G. Park, *Nat. Nanotechnol.* **9**, 927–932 (2014).
40. M. Qin *et al*., *Adv. Mater.* **32**, e2004630 (2020).
41. J. C. Hamill Jr., J. Schwartz, Y.-L. Loo, *ACS Energy Lett.* **3**, 92–97 (2018).
42. J.-W. Lee, H.-S. Kim, N.-G. Park, *Acc. Chem. Res.* **49**, 311–319 (2016).
43. A. Sharenko, C. Mackeen, L. Jewell, F. Bridges, M. F. Toney, *Chem. Mater.* **29**, 1315–1320 (2017).
44. M. V. Khenkin *et al*., *Nat. Energy* **5**, 35–49 (2020).
45. N. Li *et al*., *Science* **373**, 561–567 (2021).
46. A. Al-Ashouri *et al*., *Science* **370**, 1300–1309 (2020).
47. Z. He, shaguopohuaizhe/Schelling_simulation: fix a small bug of 1.0.0. *Zenodo* (2022); doi:10.5281/zenodo.7187438.

ACKNOWLEDGMENTS

We thank Y. Zhang, N. Li, K. Li, Y. Wu, and H. Liu (Peking University) and N. Yang, S. Ma, G. Yuan, and Y. Ma (Beijing Institute of Technology) for assistance with device fabrication; R. Si (Sun Yat-sen University) and J. Liao (Institute of High Energy Physics) for the assistance in EXAFs analysis; Z. Liu (Northwestern Polytechnical University) for the discussion on Schelling model; and Y. Wu (Auner Technology Co.) for assistance with device measurements. **Funding:** This work was supported by the National Natural Science Foundation of China (U21A20172, 21975028, 52172182, 22011540377, 62125404, U1932201, and 11705225) and Beijing Municipal Natural Science Foundation (JQ19008). **Author contribution:** Y.B. and Q.C. conceived the idea. X.Z., J.L., and Y.B. prepared the samples and carried out XPS, NMR, and optical spectroscopy characterizations. Y.B. and T.S. carried out the TOF-SIMS characterization. J.L. and H.C. carried out the DLS measurements. Y.B., Z.He, Q.C., and W.W. worked on the modification of the Schelling Model code. Yujing Li, P.A., J.Z., and J. Dong carried out the EXAFS measurement and data analysis. J.L., X.W., C.W., Z.C., J. Dou, N.X., and Y.C. carried out the other materials characterizations. M.X., Q.S., and Z.W. carried out the DFT computation. Y.B., Yiming Li, and J.S. carried out the characterization of TPC and TPV measurement and data analysis. Y.B., Z.Huang, X.Z., J.L., X.N., C.Z., H.Za., F.P., and Y.S. supported the fabrication and characterization of perovskite solar cells. H.J. and P.C. assisted in the manuscript writing. Y.B., Z.Hu., X.Z., J.L.,

X.N., H.Zh., and Q.C. carried out data analysis and wrote the manuscript. All authors discussed the results and commented on the manuscript. Y.B., Z.Hu., X.Z., J.L., and X.N. contributed equally to this work. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. The Schelling Model code is available at (*47*). **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS
science.org/doi/10.1126/science.abn3148
Materials and Methods
Figs. S1 to S40
Tables S1 to S4
References (*48*, *49*)
Movies S1 to S3

Submitted 17 November 2021; resubmitted 1 July 2022
Accepted 19 October 2022
10.1126/science.abn3148

HUMAN GENOMICS

Cross-trait assortative mating is widespread and inflates genetic correlation estimates

Richard Border[1,2,3]*, Georgios Athanasiadis[4,5], Alfonso Buil[4,6], Andrew J. Schork[4,6,7], Na Cai[8], Alexander I. Young[9,10], Thomas Werge[4,6,11], Jonathan Flint[12], Kenneth S. Kendler[13], Sriram Sankararaman[2,10,14], Andy W. Dahl[15], Noah A. Zaitlen[1,10,14]

The observation of genetic correlations between disparate human traits has been interpreted as evidence of widespread pleiotropy. Here, we introduce cross-trait assortative mating (xAM) as an alternative explanation. We observe that xAM affects many phenotypes and that phenotypic cross-mate correlation estimates are strongly associated with genetic correlation estimates (R^2 = 74%). We demonstrate that existing xAM plausibly accounts for substantial fractions of genetic correlation estimates and that previously reported genetic correlation estimates between some pairs of psychiatric disorders are congruent with xAM alone. Finally, we provide evidence for a history of xAM at the genetic level using cross-trait even/odd chromosome polygenic score correlations. Together, our results demonstrate that previous reports have likely overestimated the true genetic similarity between many phenotypes.

Methods that use summary statistics from genome-wide association studies (GWAS) to investigate genetic overlap across phenotypes have become a fundamental statistical tool across many domains of human complex trait genetics (*1*–*5*). The results of these analyses have been notable: Many trait pairs, even those with limited phenotypic similarity, display nontrivial genetic correlations [for example, 0.209 (SE = 0.042) for attention-deficit hyperactivity disorder (ADHD) and body mass index (BMI) in (*1*)]. These findings have been broadly interpreted as evidence for widespread pleiotropy across the phenome (*6*–*8*) and, in the case of psychiatric disorders, have raised concerns about the suitability of the existing nosology given evidence for shared genetic bases (*1*, *9*).

[1]Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA. [2]Department of Computer Science, University of California, Los Angeles, Los Angeles, CA 90095, USA. [3]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA. [4]Institute of Biological Psychiatry, Mental Health Center Sct Hans, Copenhagen University Hospital–Mental Health Services CPH, 2100 Copenhagen, Denmark. [5]Department of Evolutionary Biology, Ecology, and Environmental Sciences, University of Barcelona, 08028 Barcelona, Spain. [6]Globe Institute, University of Copenhagen, 1350 Copenhagen, Denmark. [7]Neurogenomics Division, The Translational Genomics Research Institute, Phoenix, AZ 85004, USA. [8]Helmholtz Pioneer Campus, Helmholtz Zentrum München, 85764 Neuherberg, Germany. [9]Anderson School of Management, University of California, Los Angeles, Los Angeles, CA 90095, USA. [10]Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA 90095, USA. [11]Department of Clinical Medicine, University of Copenhagen, 1165 Copenhagen, Denmark. [12]Center for Neurobehavioral Genetics, University of California, Los Angeles, Los Angeles, CA 90095, USA. [13]Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA 23298, USA. [14]Department of Computational Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA. [15]Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL 60637, USA.
***Corresponding author. Email: border.richard@gmail.com**

Here, we consider an overlooked source of potential bias in these findings: cross-trait assortative mating (xAM), the phenomenon whereby mates display cross-correlations across distinct traits. There are several reasons to be concerned with this potential oversight: First, the single-trait linear mixed model, which genetic correlation estimators generalize, is misspecified under single-trait assortative mating (sAM) and overestimates single-nucleotide polymorphism (SNP) heritability (*10*). Second, sAM is widespread across multiple domains for which substantial genetic correlations have been observed, including anthropometric, psychosocial, and disease traits (*1*, *7*, *8*). Third, recent work has provided genetic-level evidence for a history of sAM with respect to some of these same phenotypes (*11*). Fourth, xAM is known to generate spurious results for other marker-based inference procedures, including Mendelian randomization (*12*) and association studies (*13*).

We set out to systematically assess the impact of xAM on genetic correlation estimates, first compiling a large atlas of cross-mate correlations across a broad array of previously studied phenotypes using two large population-based samples (n = 81,394; n = 746,566). We find that these phenotypic cross-mate correlations explain a major portion of empirical marker-based genetic correlation estimates for the same trait pairs (R^2 = 74% across samples). We next demonstrate that xAM biases genetic correlation estimates and yields nontrivial estimates even among traits with uncorrelated genetic effects. We use a simulation-based approach to evaluate the extent to which empirical levels of xAM alone might plausibly explain genetic correlation estimates among previously studied traits, finding that, for many trait pairs, substantial fractions of empirical genetic correlation estimates are congruent with expectations for etiologically independent traits subject to xAM. At the same time, we observe that particular phenotype pairs, such as schizophrenia and bipolar disorders, evidence substantially larger genetic correlation estimates than can be plausibly attributed to xAM-induced artifact. Lastly, we utilize correlations between even versus odd chromosome-specific polygenic scores (PGS) to detect genetic signatures of xAM, extending a previous approach (*11*). We find that cross-trait even/odd PGS correlations mirror cross-mate phenotypic correlation patterns and, through this association, explain substantial variation in empirical genetic correlation estimates.

Results

Genetic correlation estimates mirror cross-mate phenotypic correlations

We begin by quantifying the extent to which empirical genetic correlation estimates align with cross-trait spousal correlations across a broad array of phenotypes: a set of 20 previously studied traits measured in the UK Biobank (UKB) (*14*) and a collection of six psychiatric disorder diagnoses ascertained from Danish civil registry data (*15*). We estimated cross-mate correlations for 40,697 spousal pairs within the UKB sample and 373,283 mate pairs randomly selected from the Danish population. For a pair of phenotypes Y and Z, there are three cross-mate correlation parameters: r_{yy} and

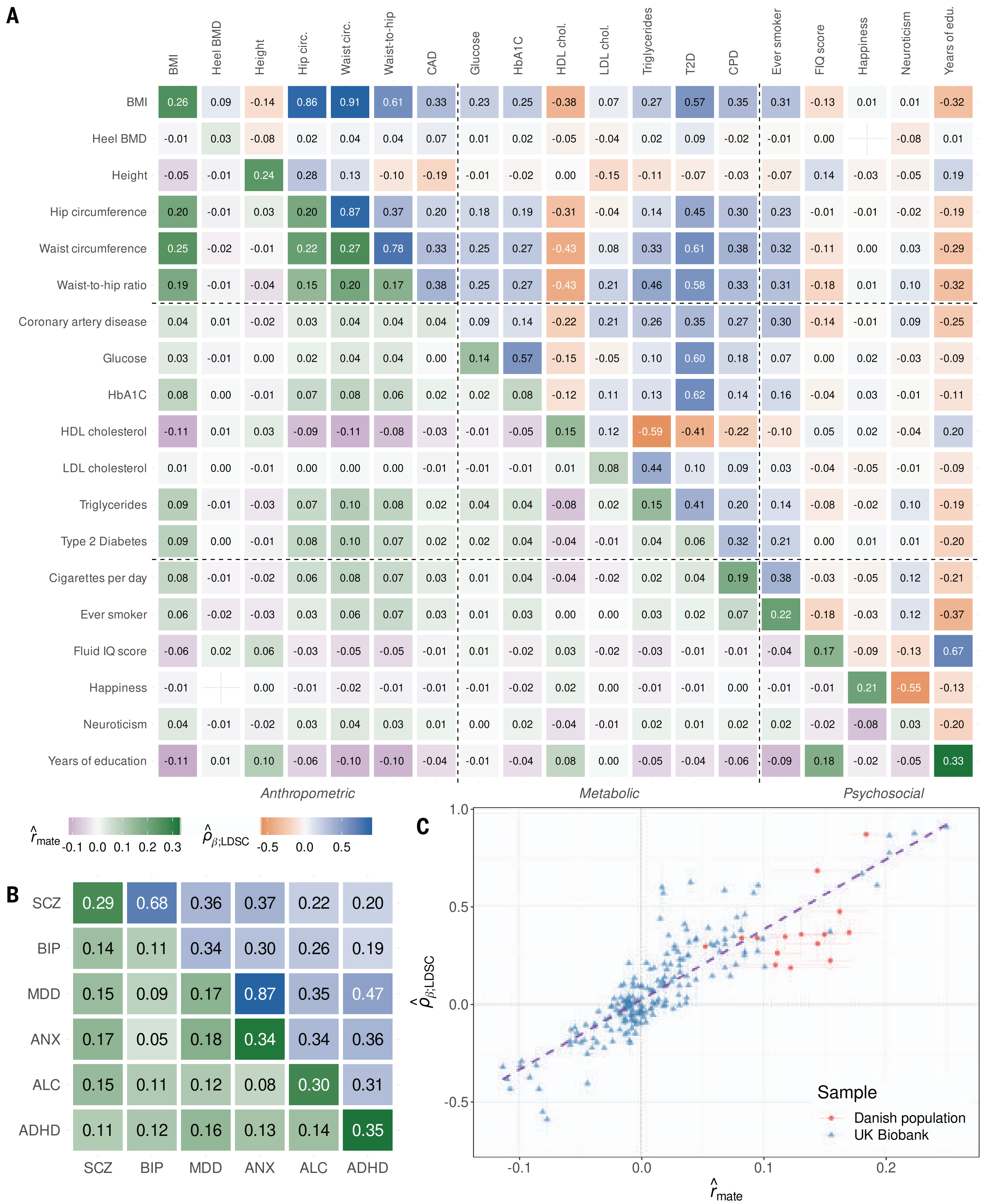


Fig. 1. Cross-mate phenotypic correlation and genetic correlation estimates. (**A**) Correlations among previously studied UK Biobank (UKB) phenotypes. Diagonal and subdiagonal heatmap entries correspond to cross-mate phenotype correlation estimates derived from 40,697 putative spouse pairs in the UKB. Superdiagonal entries correspond to empirical linkage disequilibrium score regression (LDSC) correlation estimates among unrelated European ancestry UKB participants. (**B**) Cross-mate correlation and genetic correlation estimates for psychiatric disorders. Diagonal and subdiagonal entries reflect cross-mate tetrachoric correlations among 373,283 spousal pairs sampled from the Danish population, all of which were significantly greater than zero (maximum $p = 1.69e\text{-}5$). Superdiagonal entries are previously reported LDSC correlation estimates (*23*). (**C**) Association between empirical cross-mate phenotypic correlation and genetic correlation estimates (meta-analytic $R^2 \approx 74\%$). Error lines indicate 95% confidence intervals, and the purple dashed line displays the line of best fit across all points. All numbers have been rounded to two decimal places. The model for bone mineral density (BMD) and subjective happiness failed to converge and is omitted. ADHD: attention-deficit hyperactivity disorder; ALC: alcohol use disorders; ANX: anxiety disorders; BIP: bipolar disorders; BMI: body mass index; HDL/LDL: high/low-density lipoprotein; IQ: intelligence quotient; MDD: major depressive disorder; SCZ: schizophrenia.

Table 1. Notions of genetic similarity and their relationship to genetic correlation estimators. Under random mating, score correlations and effect correlations are equal in expectation, imply the existence of pleiotropic loci, and are well captured by widely used genetic correlation estimators. Under xAM, however, substantial score correlations can arise in the absence of effect correlation or even pleiotropy, and genetic correlation estimators overestimate both ρ_β and ρ_ℓ.

Metric of genetic similarity	Relation to shared etiology and xAM
Pleiotropy is present at a particular locus when it influences both phenotypes.	Reflects shared etiology. Substantial numbers of pleiotropic loci imply that overlapping genetic variants affect both traits, though their effects may not be consistent.
Effect correlation (ρ_β) refers to the correlation between standardized genetic effects.	Reflects shared etiology. $\rho_\beta > 0$ implies that an overlapping set of variants (at pleiotropic loci) influence both traits with similar effects on average.
Score correlation (ρ_ℓ) refers to the correlation between true polygenic scores.	Reflects shared etiology, or xAM-induced population structure, or both. Roughly equal to ρ_β under random mating but larger than ρ_β under xAM owing to long-range sign-consistent LD. $\rho_\ell > 0$ does not necessarily imply biological similarity or even the existence of pleiotropic loci.
Genetic correlation estimators ($\hat\rho_\beta$), such as bivariate LD score regression, are commonly interpreted as estimates of the effect correlation.	Reflect shared etiology under random mating but produce estimates substantially greater than both ρ_β and ρ_ℓ under xAM, even when $\rho_\beta = 0$ or in the complete absence of pleiotropy.

r_{zz}, the correlations between mates on Y and Z, respectively, and r_{yz}, the cross-mate cross-trait correlation; we generically denote these quantities r_{mate} and present these estimates in the diagonal and subdiagonal entries of Fig. 1, A and B. We also compiled linkage disequilibrium score regression (LDSC) genetic correlation estimates, denoted $\hat\rho_{\beta;\text{LDSC}}$, for each pair of phenotypes, which we present in the superdiagonal entries of Fig. 1, A and B. All pairwise estimates are provided in table S1.

Cross-mate correlation structures were diverse across the trait pairs that we examined (fig. S1). Whereas cross-mate single-trait and cross-mate cross-trait correlations were similar for some trait pairs (for example, $\hat r_{yy}$ = 0.26, $\hat r_{zz}$ = 0.20, and $\hat r_{yz}$ = 0.20 for BMI and hip circumference), these quantities were of opposing signs for others (for example, $\hat r_{yy} = 0.33$ and $\hat r_{zz} = 0.22$ versus $\hat r_{yz} = -0.09$ for years of education and regular smoking). In general, cross-mate correlation structures were not consistent with sAM alone. When the cross-mate correlation r_{zz} for a secondary trait Z is fully mediated through sAM on Y, we expect that $\hat r_{zz} \approx \hat r_{yy}\hat s_{yz}^2$; this model fit the data poorly (fig. S2).

Estimates of $\hat\rho_{\beta;\text{LDSC}}$ were strongly associated with r_{mate} estimates across both samples [Fig. 1C; meta-analytic R^2 = 74.32%; 95% confidence interval (CI): 67.02%–81.62% in a linear model; R^2 = 76.69%; 95% CI: 73.94%–79.45% in a Bayesian model accounting for heteroskedasticity and estimation error]. The regression slope did not significantly differ across the UKB and psychiatric phenotypes in either model (for example, p = 0.16 for a sample-by-r_{mate} interaction term in the linear model). The strength of this association largely persisted when excluding trait pairs with large genetic correlation estimates: Considering only trait pairs with estimated genetic correlations below 0.50 in magnitude yielded R^2 = 70.94%; further restricting to those below 0.30 in magnitude, yielded R^2 = 67.88%. This suggests that the observed association does not merely reflect sAM on genetically homogeneous factors.

Defining genetic correlation

Having established that a large degree of the variance in genetic correlation estimates can be predicted from phenotypic mating correlations, we now provide theoretical intuition as to why this might occur [see supplementary text for further details (*16*)]. We start by defining three distinct notions of genetic similarity between phenotypes. These definitions are summarized in Table 1.

We consider a pair of phenotypes Y, Z, with heritable components ℓ_y, ℓ_z reflecting the additive effects of m standardized haploid variants $X_1, ..., X_m$ with phenotype-specific effect vectors $\boldsymbol\beta_y$, $\boldsymbol\beta_z$. For simplicity, we assume that causal variants are initially unlinked and that both phenotypes have unit variance under random mating (panmixis), such that the panmictic heritabilities are $h^2_{y;\text{pan}} = \boldsymbol\beta_y^\top\boldsymbol\beta_y$ and $h^2_{z;\text{pan}} = \boldsymbol\beta_z^\top\boldsymbol\beta_z$.

Pleiotropy is present when a locus influences two or more phenotypes. Thus, locus X_i is pleiotropic with respect to Y and Z when both $\beta_{y;i} \neq 0$ and $\beta_{z;i} \neq 0$, though these effects might differ substantially in magnitude or direction. By contrast, the correlation between effects, which we refer to as the effect correlation ρ_β, indexes the similarity of variant effects on two phenotypes:

$$\rho_\beta = \text{cor}(\boldsymbol\beta_y, \boldsymbol\beta_z)$$

where the $\boldsymbol\beta$ vectors include all variants, causal or otherwise, and thus may contain elements equal to zero. A value of $\rho_\beta > 0$ implies both the existence of pleiotropic loci and that such loci have similar effects on average, and we term a pair of traits genetically orthogonal when $\rho_\beta = 0$. Effect correlation is distinct from the classical definition of genetic correlation as the correlation between the heritable components of two traits (*17*), which we refer to as the score correlation:

$$\rho_\ell = \text{cor}(\ell_y, \ell_z)$$

as it reflects the correlation between the true PGS.

Within the standard linear mixed model framework, ρ_ℓ and ρ_β are equivalent and hence seldom discussed as separate quantities [though genetic correlation estimates are commonly interpreted as estimates of ρ_β (*4*, *18*, *19*)]. However, traits with uncorrelated effects can unintuitively have correlated PGS. Under xAM, all causal variants affecting trait Y become correlated with all causal variants affecting trait Z, and these correlations are directionally consistent with their respective effects [see supplementary text (*16*)]. As we will demonstrate in the following section, this results in nonzero score correlations in the direction of the cross-mate cross-trait phenotypic correlation, even for genetically orthogonal traits.

The impact of xAM in simulations

We ran a series of forward-time simulations using realistic genotype data to investigate the impact of xAM on multiple measures of genetic correlation. At each generation, individuals (consisting of a set of genotypes together with two phenotypes) were matched to achieve

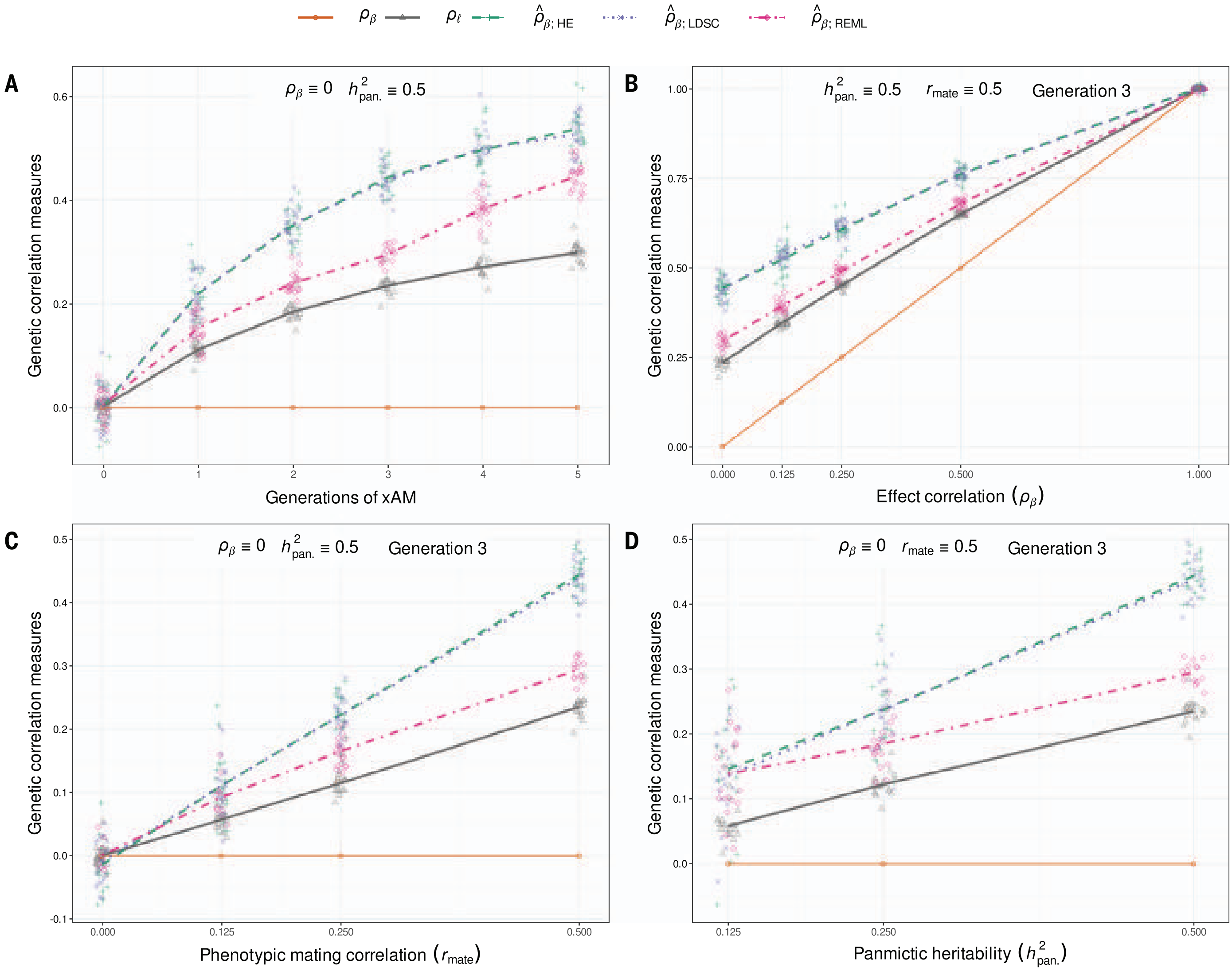


Fig. 2. Impact of xAM on genetic correlation estimates in forward-time simulations. Score correlation (ρ_ℓ) and genetic correlation estimates ($\hat{\rho}_\beta$) for two phenotypes with true effect correlation ρ_β, panmictic heritabilities h^2_{pan}, and all cross-mate correlations set to r_{mate}. (**A**) xAM increases the true score correlation among genetically orthogonal phenotypes. HE, LDSC, and REML estimators all further overestimate ρ_β, and the magnitude of this bias increases over subsequent generations. (**B**) After three generations of xAM, $\hat{\rho}_\beta$ estimates are upwardly biased for genetically distinct phenotypes. (**C**) The impact of three generations of xAM increases with the cross-mate correlation. (**D**) The impact of three generations of xAM increases with the panmictic heritabilities.

target cross-mate correlation parameters, after which we estimated genetic correlations ($\hat{\rho}_\beta$) using LDSC [denoted $\hat{\rho}_{\beta;LDSC}$ (*6*)], Haseman-Elston regression [HE, denoted $\hat{\rho}_{\beta;HE}$ (*20*)], and residual maximum likelihood [REML; $\hat{\rho}_{\beta;REML}$ (*18*)]. We also computed true score correlations (ρ_ℓ), which is possible when the true genetic effects are known. We performed sensitivity analyses to confirm that our results did not depend on simulation parameters, including the number of causal variants (fig. S3), mate selection algorithm (fig. S4), recombination scheme (fig. S5), and whether causal variants with orthogonal genetic effects arose on overlapping loci (fig. S6). We additionally investigated the impact of xAM on GWAS effect estimates and GWAS-based methods for identifying pleiotropic SNPs (figs. S7 to S10), genetic correlation estimates for binary phenotypes subject to misdiagnosis (fig. S11), partitioned genetic correlation estimates (fig. S12), and genetic covariance estimators (fig. S13).

xAM induces nonzero score correlations among genetically orthogonal traits

We confirmed that xAM induces substantial score correlations across a broad array of simulation parameters. This is perhaps most pronounced for traits with orthogonal effects: Fig. 2A demonstrates the increase in the true score correlation across multiple generations of xAM for a pair of traits with $\rho_\beta = 0$, $r_{mate} = 0.5$, and $h^2_{pan} = 0.5$. Across simulation replicates, the average score correlation was 0.11 after a single generation of xAM, which increased to 0.24 after three generations.

Notably, this increase in score correlation induced by xAM does not represent bias: The population-level correlation between the heritable components of the phenotypes truly does increase under xAM. However, as we demonstrate next, genetic correlation estimators become misspecified under xAM and yield biased estimates. Still, even unbiased estimates of

score correlation can be driven by either shared biology, or xAM, or both, further complicating interpretation (Table 1).

Effect correlation estimates are biased upward

For genetically orthogonal traits, after a single generation of xAM, the REML estimator yielded $\hat{\rho}_\beta$ = 0.15 and the HE and LDSC estimators, which are closely related (*21*, *22*), both yielded average estimates of $\hat{\rho}_\beta$ = 0.21, all of which were greater than the true value of ρ_β = 0.00. After three generations of xAM, this upward bias became more pronounced, with REML and LDSC yielding estimates of $\hat{\rho}_\beta$ = 0.30 and $\hat{\rho}_\beta$ = 0.44, respectively.

The quantities ρ_ℓ, and $\hat{\rho}_\beta$ are monotonically related to ρ_β, r_{mate}, and h^2_{pan}

Many trait pairs subject to xAM will truly have correlated genetic effects. Figure 2B illustrates the relationship between ρ_β, ρ_ℓ, and $\hat{\rho}_\beta$ for two traits with h^2_{pan}= 0.5. Excepting the case of ρ_β = 1.0 (genetically identical phenotypes), results remained consistent with the genetically orthogonal case: ρ_β was lower than ρ_ℓ, which was in turn lower than the upwardly biased $\hat{\rho}_\beta$ estimates provided by REML, HE, and LDSC. For example, when ρ_β = 0.25, LDSC yielded $\hat{\rho}_\beta$ = 0.62 after three generations of xAM. We note that whereas the true effect correlation varies in Fig. 2B, the cross-mate correlations remain fixed, demonstrating that the potential for xAM-induced bias is present even when cross-mate cross-trait correlations partially reflect shared genetic bases. The impact of xAM on both ρ_ℓ and $\hat{\rho}_\beta$ was greater for traits under stronger xAM (Fig. 2C) and for traits with greater heritabilities (Fig. 2D).

xAM biases annotation- and locus-level analyses

Partitioned genetic correlation estimators evidenced similar biases as genome-wide estimators under xAM, even when supplied with annotations directly relevant to bivariate genetic architecture. Further, this bias was greatest at regions relevant to only one of the two phenotypes (fig. S12).

In association studies, GWAS effect estimates for SNPs causal for the focal trait were biased upward in magnitude, whereas those causal for a secondary, unrelated trait under xAM with the first were biased toward their effects on that trait (figs. S7 to S9). These biases were asymptotically non-negligible (fig. S9). As a result, xAM increased the likelihood of rejecting the null hypothesis of no association at all SNPs causal for either trait, increasing both statistical power and false-positive rates (fig. S10). Eventually, all variants affecting a secondary phenotype subject to xAM with the GWAS phenotype will reach genome-wide significance as sample size becomes large. However, spurious effect estimates will remain attenuated (figs. S7 to S9), implying that methods for identifying cross-trait heterogeneity in GWAS estimates may have the potential to differentiate trait-specific signal from xAM-induced artifacts.

xAM alone can plausibly explain substantial variance in empirical genetic correlation estimates

We next sought to quantify the extent to which empirical $\hat{\rho}_\beta$ estimates for previously studied trait pairs could be explained by xAM alone, assuming genetic orthogonality. We proceeded with a simulation-based approach, which we present for the UKB and psychiatric phenotypes in turn.

Within each sample, we identified mate pairs and estimated phenotypic cross-mate correlations. We then used these estimates, together with empirical heritability estimates, as inputs to a forward-time simulation where separate, nonoverlapping collections of causal variants were assigned to each phenotype, such that ρ_β= 0.0. At each generation, we estimated the effect correlation, ρ_β, using method of moments. These projected effect correlation estimates, which we denote $\hat{\rho}_{\text{xAM}}$, can be interpreted as expected LDSC genetic correlation estimates for genetically orthogonal traits under xAM consistent with empirical spousal correlations.

We next compared $\hat{\rho}_{\text{xAM}}$ to empirical LDSC estimates derived in real data, which we denote $\hat{\rho}_{\text{emp}}$. To simplify discussion, we define the ratio

$$\hat{\gamma} = \hat{\rho}_{\text{xAM}}/\hat{\rho}_{\text{emp}}$$

which measures the projected LDSC effect correlation estimate due to xAM-induced artifact relative to the empirical LDSC effect correlation estimate for a given phenotype pair (Fig. 3A).

Expected effect correlation estimates for UKB phenotypes in the absence of pleiotropy

We restricted our attention to 132 (of 190 possible) pairs of UKB phenotypes with nominally significant ($p < 0.05$) LDSC genetic correlation estimates (table S1). We first obtained pedigree-based heritability estimates for each of the traits of interest from the literature, using estimates derived in demographically comparable samples (table S2). Together with the phenotypic mating correlations (Fig. 1A), these comprised inputs to forward time simulations that we used to compute $\hat{\rho}_{\text{xAM}}$.

Across 132 trait pairs, 42 evidenced $\hat{\gamma}$ values significantly greater than zero (their 95% credible intervals did not include zero) after a single generation of xAM, which increased to 74 trait pairs after three generations. Across all trait pairs (including those not significantly different from zero), the inverse variance weighted average $\hat{\gamma}$ estimate was 0.25 (SE = 0.005). Figure 3B presents the first 20 pairs in descending order of $\hat{\gamma}$, and Fig. 3C presents the raw projected and empirical effect correlation estimates across all 132 pairs (see table S3 and figs. S14 and S15 for detailed results spanning five generations of xAM). Finally, Fig. 3D displays average $\hat{\gamma}$ values within and between qualitative phenotypic domains.

Expected effect correlation estimates among psychiatric disorders in the absence of pleiotropy

We next estimated $\hat{\rho}_{\text{xAM}}$ for a collection of six psychiatric disorders, using correlations estimated in spousal pairs randomly selected from the Danish population (table S4). We then compared these projections to the LDSC genetic correlation estimates reported by Grotzinger and colleagues (*23*).

Across all pairwise combinations of disorders, we observed an average ratio of $\hat{\gamma}$ = 0.29 (SE = 0.016; Fig. 4, A and B) after five generations of xAM. Some trait pairs evidenced considerably greater empirical genetic correlation estimates than might be explained by xAM alone (for anxiety disorders and major depression, $\hat{\gamma}$ = 0.21; 95% CI: 0.17–0.25), whereas for other pairs this discrepancy was modest (for alcohol use disorder and schizophrenia, $\hat{\gamma}$ = 0.83; 95% CI: 0.59–1.24; see table S5 and fig. S16 for complete results).

xAM exacerbates bias due to misdiagnosis

After additional simulations demonstrated that xAM further inflates genetic correlation estimates in the context of diagnostic errors (Fig. S11), we extended our method for estimating $\hat{\rho}_{\text{xAM}}$ to incorporate misdiagnosis. Results were heterogeneous across disorder pairs (fig. S17). For example, whereas moderate rates of diagnostic errors (5%), together with three generations of xAM, yielded genetic correlation estimates for ADHD and major depression on par with published estimates ($\hat{\gamma}$ = 0.97; 95% CI: 0.73–1.22), substantial diagnostic errors (15%) after five generations of xAM yielded estimates well below previously published estimates for bipolar disorders and major depression ($\hat{\gamma}$ = 0.37; 95% CI: 0.12–0.62). Figure 4C highlights the potential impacts of xAM and diagnostic errors on four selected trait pairs, and fig. S17 presents results for all pairs.

Genetic evidence for xAM recapitulates empirical cross-mate correlations

Cross-mate phenotypic correlation estimates (r_{mate}) explained substantial variance in the cross-chromosome even/odd PGS correlations in a linear model ($\hat{\rho}_{\ell;\text{eo}}$; R^2 = 47.66%; Fig. 5A). This association, which is congruent with expectations under phenotypically mediated xAM [supplementary text (*16*)], persisted when

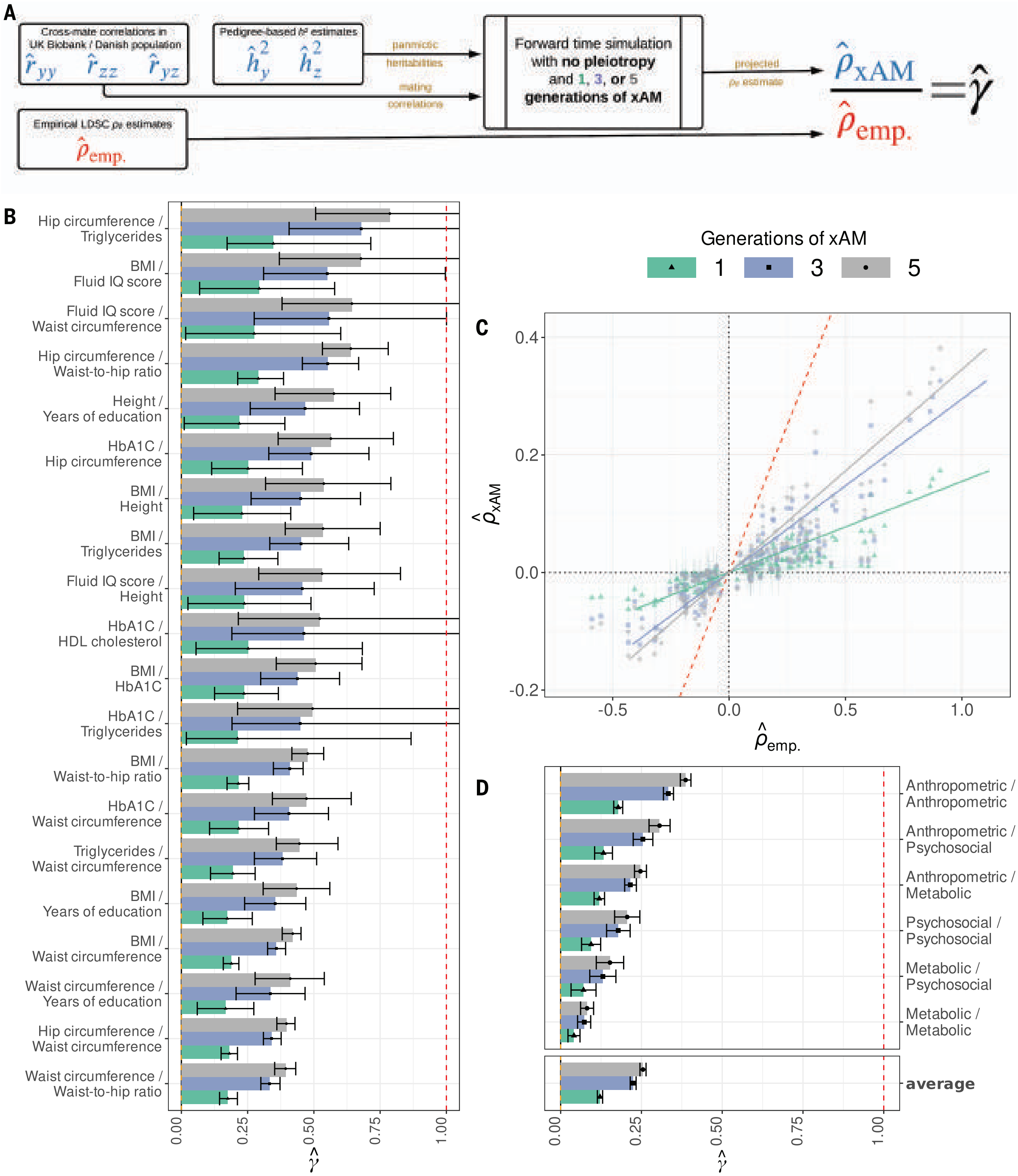


Fig. 3. Empirical and expected genetic correlations among UK Biobank phenotypes. (**A**) We computed the expected LDSC genetic correlation estimate in the absence of pleiotropy and after a given number of generations of xAM ($\hat{\rho}_{xAM}$), which we compared to empirical LDSC estimates ($\hat{\rho}_{emp}$) to obtain the ratio $\hat{\gamma} = \hat{\rho}_{xAM}/\hat{\rho}_{emp}$. (**B**) The top 20 $\hat{\gamma}$ estimates across previously studied UKB phenotype pairs with nominally significant ($p < 0.05$) $\hat{\rho}_{emp}$ values. (**C**) Projected versus empirical LDSC estimates for all UKB phenotypes with nominally significant genetic correlation estimates. (**D**) Inverse variance weighted average $\hat{\gamma}$ estimates within and between qualitatively similar phenotypic domains. Error bars throughout represent 95% credible intervals.

accounting for measurement error and heteroskedasticity, and across PGS p-value thresholds (fig. S18).

Additionally, cross-trait even/odd chromosome PGS correlations were positively associated with empirical LDSC genetic correlation estimates ($\hat{\rho}_{\beta;LDSC}$; $R^2 = 34.81\%$; Fig. 5B). This is consistent with the hypothesis that empirical effect correlation estimates are capturing additional structure beyond the signatures of biological overlap. Further, regressing $\hat{\rho}_{\beta;LDSC}$ on $\hat{\rho}_{\ell;eo}$ and $\hat{r}_{yz}$ simultaneously revealed that the association between $\hat{\rho}_{\beta;LDSC}$ and $\hat{\rho}_{\ell;eo}$ is mediated via $\hat{r}_{yz}$ ($\Delta R^2 < 0.001$; partial effect $p = 0.48$ for $\hat{\rho}_{\ell;eo}$ versus $p <$ 5e-8 for $\hat{r}_{yz}$). Thus,

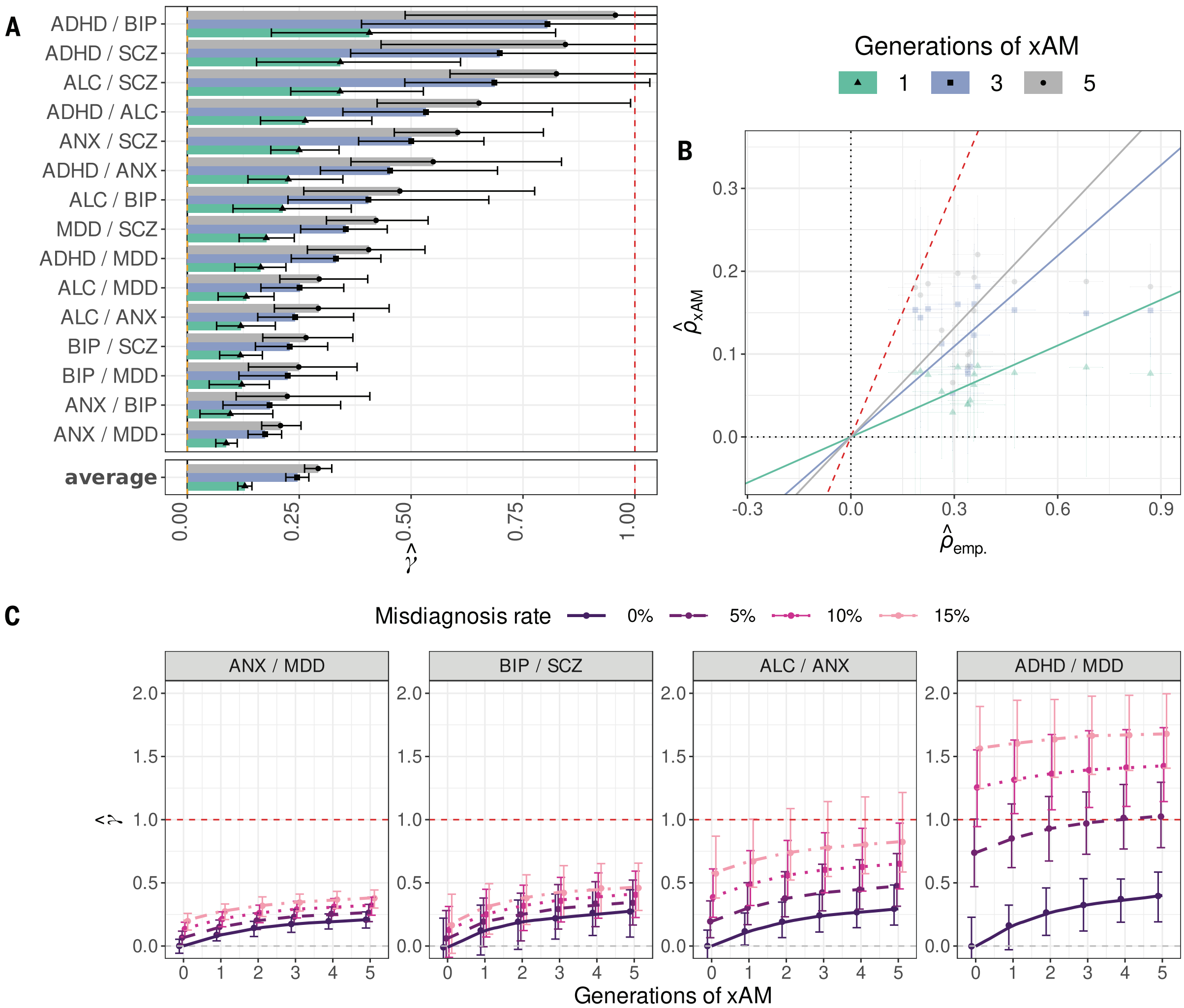


Fig. 4. Empirical and expected genetic correlations among psychiatric phenotypes. (**A**) Ratios ($\hat{\gamma} = \hat{\rho}_{xAM}/\hat{\rho}_{emp}$) of projected LDSC genetic correlation estimates under xAM alone relative to empirical genetic correlation estimates (*23*) for six psychiatric disorders with 95% credible intervals. (**B**) Projected versus empirical LDSC estimates across psychiatric phenotype pairs. (**C**) The potential combined impacts of bidirectional errors in diagnosis and xAM on genetic correlation estimates for selected psychiatric disorder pairs. The red dashed line corresponds to $\hat{\gamma} = 1$ across all panels.

alternative sources of structure independent from xAM do not appear to explain the positive association between $\hat{\rho}_{\ell;eo}$ and $\hat{\rho}_{\beta;LDSC}$.

Discussion

Nonzero effect correlation estimates have been widely interpreted as evidence for overlapping genetic bases. It is therefore surprising that substantial variation in genetic correlation estimates can be explained by cross-mate phenotypic correlations. Given the strength of this association, the consequences of the random mating assumption implicit in all commonly used genetic correlation estimators warrant critical attention.

Our results show that cross-mate phenotypic correlations among many pairs of phenotypes are strong enough that one or more generations of xAM would substantially inflate genetic correlation estimates. Further, the correlation structures of even/odd chromosome PGS coincide with expectations after one or more generations of xAM. We conclude that xAM comprises a source of systematic bias in the study of genetic similarity across complex traits, one that subverts the widespread interpretation of genetic correlation as a direct index of biological similarity.

Our findings mirror recent results regarding the potential impacts of assortative mating across other areas of statistical genetics, including marker-based heritability estimation (*10*) and Mendelian randomization (*12*). Our results also complicate the interpretation of a number of multivariate analytic frameworks. For example, genomic structural equation modeling (*24*), which takes marker-based genetic correlation estimates as inputs, will propagate xAM-induced biases. This does not mean such methods are fundamentally flawed, but instead demonstrates the importance of developing unbiased effect correlation estimators given the centrality of genetic correlation estimates in modern statistical genetics.

With this in mind, we comment on potential approaches to disentangling true effect correlation from xAM-induced artifact. First, family-based designs for addressing xAM (*25*) are increasingly being applied to molecular genetic data with promising results (*26*). Second, we conjecture that approaches aimed at characterizing effect heterogeneity across multiple phenotypes may provide a viable means for identifying trait-specific loci: Although all trait-specific loci for either of two traits will achieve genome-wide significance in large sample GWAS of either trait, effect estimates will remain substantially larger at causal loci. Finally, we propose that directly modeling the dependence between genotypes and their effects will allow the differentiation of effect correlations and score correlations in samples of unrelated individuals.

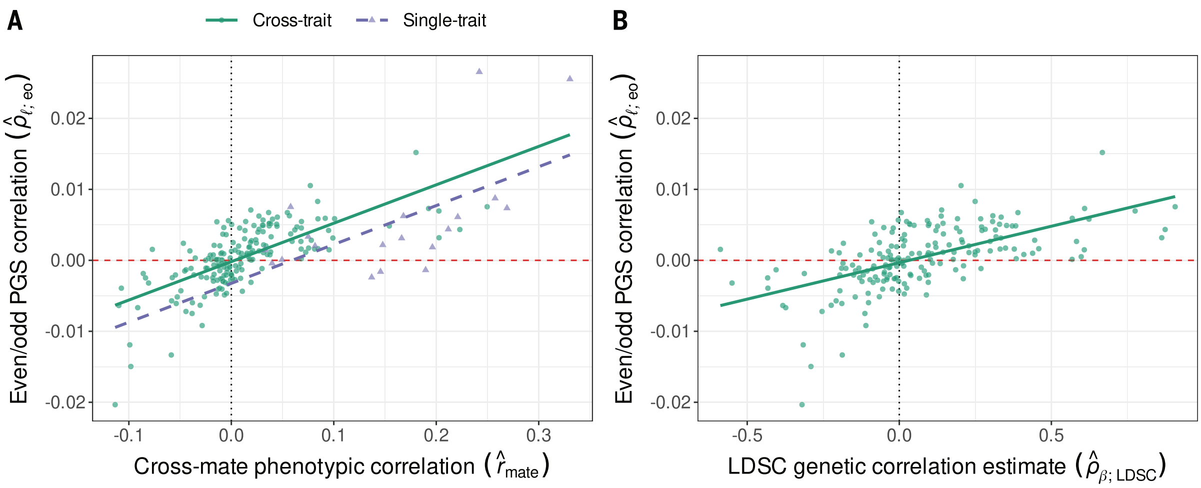


Fig. 5. Genetic level evidence consistent with xAM in the UK Biobank. (**A**) Correlation between even and odd chromosome-specific polygenic scores (PGS) as a function of the cross-mate phenotypic correlation. For a single trait, the vertical axis reflects the correlation between even and odd chromosome scores $\hat{\ell}_{\mathrm{even}}$, $\hat{\ell}_{\mathrm{odd}}$ and the horizontal axis reflects the cross-mate correlation. For a pair of traits Y, Z, the vertical axis reflects a single parameter to which the correlations between $\hat{\ell}_{y;\mathrm{even}}$, $\hat{\ell}_{z;\mathrm{odd}}$ and between $\hat{\ell}_{y;\mathrm{odd}}$, $\hat{\ell}_{z;\mathrm{even}}$ are both constrained, and the horizontal axis reflects the cross-mate cross-trait correlation. (**B**) Cross-trait even/odd PGS correlations as a function of empirical LD score regression genetic correlation estimates.

There are several limitations to the current investigation. Foremost among these are the numerous assumptions about population dynamics required to model xAM, including but not limited to panmictic heritabilities, stability of cross-mate phenotypic correlation structures over succeeding generations, stability and transmissability of environmental factors, and the extent to which mating patterns reflect social versus genetic homogamy. We proceeded under the tractable dynamical framework of two additive phenotypes subject to primary-phenotypic xAM with constant cross-mate correlations, stable nonheritable sources of variation, and no vertical transmission. Though each of these assumptions is likely untenable for particular trait pairs, thereby compromising the accuracy of our projections, we hypothesize that the qualitative phenomenon whereby xAM inflates genetic correlation estimates will persist for many traits. Nonetheless, we caution that these projections are contingent upon multiple consequential decisions. Constructing a generative model that reconciles the association between empirical mating patterns and genetic correlation estimates is an ill-posed inverse problem for which there are multiple solutions and of which we have only explored a subset. At the same time, existing methods are only able to sidestep these decisions by making the strong (and often incorrect) assumption that mating is random.

Lastly, we remark that xAM is, in essence, a form of population structure not captured by conventional principal-component– or mixed-model–based correction. Given the increasing evidence that existing methods fail to completely address structural factors, even in ostensibly ancestrally homogeneous groups (*27*), a broader characterization of population structure and methods for addressing such structure will likely be necessary to generate results that are maximally clinically relevant and can be applied equitably.

REFERENCES AND NOTES

1. V. Anttila *et al.*, *Science* **360**, eaap8757 (2018).
2. P. H. Lee, Y. A. Feng, J. W. Smoller, *Biol. Psychiatry* **89**, 20–31 (2021).
3. O. Weissbrod, J. Flint, S. Rosset, *Am. J. Hum. Genet.* **103**, 89–99 (2018).
4. B. K. Bulik-Sullivan *et al.*, *Nat. Genet.* **47**, 291–295 (2015).
5. Y. Wu *et al.*, *Nat. Commun.* **10**, 1891 (2019).
6. B. Bulik-Sullivan *et al.*, *Nat. Genet.* **47**, 1236–1241 (2015).
7. J. Zheng *et al.*, *Bioinformatics* **33**, 272–279 (2017).
8. W. van Rheenen, W. J. Peyrot, A. J. Schork, S. H. Lee, N. R. Wray, *Nat. Rev. Genet.* **20**, 567–581 (2019).
9. P. H. Lee *et al. Cell* **179**, 1469–1482.e11 (2019).
10. R. Border *et al.*, *Nat. Commun.* **13**, 660 (2022).
11. L. Yengo *et al.*, *Nat. Hum. Behav.* **2**, 948–954 (2018).
12. F. P. Hartwig, N. M. Davies, G. Davey Smith, *Genet. Epidemiol.* **42**, 608–620 (2018).
13. A. Kong *et al.*, *Science* **359**, 424–428 (2018).
14. C. Bycroft *et al.*, *Nature* **562**, 203–209 (2018).
15. O. Plana-Ripoll *et al.*, *JAMA Psychiatry* **76**, 259–270 (2019).
16. Materials and methods are available as supplementary materials.
17. D. S. Falconer, *Introduction to Quantitative Genetics* (Oliver and Boyd, 1960).
18. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, *Am. J. Hum. Genet.* **88**, 76–82 (2011).
19. P.-R. Loh, G. Kichaev, S. Gazal, A. P. Schoech, A. L. Price, *Nat. Genet.* **50**, 906–908 (2018).
20. J. K. Haseman, R. C. Elston, *Behav. Genet.* **2**, 3–19 (1972).
21. B. Bulik-Sullivan, Relationship between LD Score and Haseman-Elston Regression. bioRxiv [Preprint]. 20 April 2015. https://doi.org/10.1101/018283.
22. X. Zhou, *Ann. Appl. Stat.* **11**, 2027–2051 (2017).
23. A. D. Grotzinger *et al.*, *Nat. Genet.* **54**, 548–559 (2022).
24. A. D. Grotzinger *et al.*, *Nat. Hum. Behav.* **3**, 513–525 (2019).
25. M. C. Keller *et al.*, *PLOS Genet.* **9**, e1003451 (2013).
26. L. J. Howe *et al.*, *Nat. Genet.* **54**, 581–592 (2022).
27. J. J. Berg *et al.*, *eLife* **8**, e39725 (2019).
28. R. Border, xAM_and_gen_corr, Zenodo (2022). https://doi.org/10.5281/zenodo.7065696.
29. Statistics Denmark, Data for research; https://www.dst.dk/en/TilSalg/Forskningsservice.

ACKNOWLEDGMENTS

We thank the editors and referees as well as A. L. Price and M. C. Keller for their thorough and helpful feedback. **Funding:** Chan Zuckerberg Initiative grant CZF2019-002449 (N.A.Z., R.B.), NIH grant R01AG042568 (A.I.Y.), NIH grant R01CA227237 (N.A.Z., R.B.), NIH grant R01CA227466 (N.A.Z., R.B.), NIH grant R01ES029929 (N.A.Z., R.B.), NIH grant R01GM142112 (N.A.Z., R.B.), NIH grant R01HG006399 (N.A.Z., R.B.), NIH grant R01HG011345 (N.A.Z., R.B.), NIH grant R01HL151152 (N.A.Z., R.B.), NIH grant R01HL155024 (N.A.Z., R.B.), NIH grant R01MH125252 (N.A.Z., R.B.), NIH grant R01MH122688 (N.A.Z., R.B.), NIH grant R01MH130581 (A.W.D., J.F., K.S.K., N.A.Z., N.C.), NIH grant R24AG065184 (A.I.Y.), NIH grant R35GM125055 (S.S.), NIH grant R35GM133531 (N.A.Z., R.B.), NIH grant T32NS048004 (R.B.), NIH grant U01MH126798 (N.A.Z., R.B.), NIH grant U01HG009080 (N.A.Z., R.B.), National Science Foundation award CAREER-1943497 (S.S.), Open Philanthropy grant 010623-00001 (A.I.Y.), Wellcome Trust grant 200176/Z/15/Z (J.F., K.S.K.). **Author contributions:** Conceptualization: A.W.D., J.F., K.S.K., N.A.Z., R.B., T.W. Methodology: A.B., A.I.Y., A.J.S., A.W.D., G.A., N.A.Z., N.C., R.B., S.S., T.W. Investigation: G.A., R.B. Visualization: R.B. Funding acquisition: A.W.D., J.F., K.S.K., N.A.Z., N.C., S.S. Software: R.B. Supervision: A.W.D., J.F., K.S.K., N.A.Z., S.S. Writing – original draft: A.W.D., N.A.Z., R.B. Writing – review and editing: A.I.Y., A.J.S., A.W.D., J.F., N.A.Z., R.B. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Cross-mate correlation estimates and all other results are provided in tables S1 to S6 (*16*). Code used to perform simulations is publicly available (*28*). UK Biobank data were accessed under application 33127 and are available through the UK Biobank Access Management System http://amsportal.ukbiobank.ac.uk/. Danish data are stored in a national high performance computing facility in Denmark. Access to data may be granted upon application to Statistics Denmark in accordance with Danish Law as detailed online (*29*). **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/content/page/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo2059
Materials and Methods
Supplementary Text
Figs. S1 to S20
Tables S1 to S7
References (*30–46*)
MDAR Reproducibility Checklist

Submitted 21 January 2022; accepted 28 September 2022
10.1126/science.abo2059

PLANT SCIENCE

Hydraulic flux–responsive hormone redistribution determines root branching

Poonam Mehra[1]*, Bipin K. Pandey[1], Dalia Melebari[1], Jason Banda[1], Nicola Leftley[1], Valentin Couvreur[2], James Rowe[3], Moran Anfang[4], Hugues De Gernier[5,6], Emily Morris[1]†, Craig J. Sturrock[1], Sacha J. Mooney[1], Ranjan Swarup[1], Christine Faulkner[7], Tom Beeckman[5,6], Rishikesh P. Bhalerao[8], Eilon Shani[4], Alexander M. Jones[3], Ian C. Dodd[9], Robert E. Sharp[10], Ari Sadanandom[11], Xavier Draye[2], Malcolm J. Bennett[1]*

Plant roots exhibit plasticity in their branching patterns to forage efficiently for heterogeneously distributed resources, such as soil water. The xerobranching response represses lateral root formation when roots lose contact with water. Here, we show that xerobranching is regulated by radial movement of the phloem-derived hormone abscisic acid, which disrupts intercellular communication between inner and outer cell layers through plasmodesmata. Closure of these intercellular pores disrupts the inward movement of the hormone signal auxin, blocking lateral root branching. Once root tips regain contact with moisture, the abscisic acid response rapidly attenuates. Our study reveals how roots adapt their branching pattern to heterogeneous soil water conditions by linking changes in hydraulic flux with dynamic hormone redistribution.

Root branching determines foraging capacity of crop plants (*1*, *2*). Roots have evolved plasticity in their branching patterns to maximize access to soil water resources, which are often distributed heterogeneously (*3*, *4*). Root branching is affected both by extreme weather (such as drought or flooding) (*5*, *6*) and by transient or local spatial differences in soil moisture (*7*, *8*). Negative effects of water scarcity on modern agriculture will be exacerbated as climate change affects hydrological cycles and soil water resources (*9*, *10*). Increased resilience in the face of climate change will require better insight into how plant roots sense and adapt to fluctuating water availability.

Dissecting roots' water sensing using xerobranching

Xerobranching (*7*) provides an experimental model to study root adaptive responses to transient water stress. A xerobranching response is triggered when growing root tips temporarily lose contact with moist soil (e.g., in an air gap), which causes branching to temporarily cease until roots reenter moist soil (Fig. 1A). To discover the mechanistic basis of xerobranching, an agar-based xerobranching bioassay was developed (fig. S1) that phenocopies the original x-ray computed tomography (CT) soil-based bioassay (Fig. 1A) in experiments with *Arabidopsis* and tomato seedlings. As levels of the abiotic stress signal abscisic acid (ABA) increase in root tips during transient water stress (*7*), we tested whether tomato ABA biosynthesis mutants (*flacca* and *notabilis*) (*11*, *12*) are disrupted in xerobranching response. Both soil- and agar-based bioassays revealed that, unlike wild-type (WT) controls, primary roots of both *flacca* and *notabilis* continue to branch when growing across an air gap (Fig. 1, A to D, and fig. S2). Similarly, the ABA-deficient maize mutant *vp14* (*13*) disrupted xerobranching in both paper-based and soil-based bioassays (figs. S3 and S4). Hence, ABA is a widely conserved regulator that represses lateral root development during a xerobranching response in these monocot and eudicot plant models.

From which tissue(s) does ABA originate to trigger a xerobranching response? ABA biosynthesis genes, such as *ABA2*, are expressed in phloem-related root vascular tissues (*14*). *Arabidopsis aba2-1* mutant roots exhibit a xerobranching defect, which can be rescued by expressing WT *ABA2* using the phloem companion cell–expressed *SUC2* promoter (Fig. 1, E to G, and fig. S5A). A *SUC2:YFP-ABA2* translation fusion confirmed that the ABA2 protein can act in a phloem companion cell–autonomous manner (fig. S5B). Phloem provides water for growing roots in the temporary absence of an external supply (*15*). Hydraulic modeling (*16*, *17*) predicts that radial water fluxes within root tissues change direction from predominately inward to outward after a xerobranching stimulus (Fig. 1H). Simulations reveal that an inward water flux dominates under normal externally hydrated conditions, and phloem-derived water is limited to close to companion cells (Fig. 1H). However, after a xerobranching stimulus (when the external water source is temporarily lost), root-tip tissues rely on an outward flux of phloem-derived water to maintain growth, requiring ~8 hours for outer root tissues to receive phloem water (Fig. 1H). Phloem-derived ABA is likely to travel with this outward flow of water during a xerobranching response.

ABA moves outward after a xerobranching stimulus

To visualize whether, when, and where ABA movement occurs after a xerobranching stimulus, we used a sensitive Förster resonance energy transfer (FRET)–based hormone biosensor (*18*), nlsABACUS2, which displays increased fluorescence emission ratio upon ABA binding (fig. S6). Seedlings expressing a nuclear-localized nlsABACUS2 biosensor revealed dynamic changes in ABA distribution and levels during our agar-based xerobranching bioassay (Fig. 1, I to K, and figs. S7 and S8). Spatially, nlsABACUS2 accumulates in epidermal cells in the basal meristem and elongation zones (Fig. 1J). Quantifying temporal changes in emission ratios revealed that ABA levels initially increase in root epidermal tissues >12 hours after a xerobranching stimulus. Considering the sensitivity of the biosensor, these results are also consistent with our hydraulic simulations that predict the outward flux of phloem-derived water after ~8 hours of xerobranching stimulus (Fig. 1H). Once root tips reconnect with the agar (recovery phase), ABA levels fall rapidly in epidermal cells (Fig. 1, J and K, and fig. S8).

To directly visualize whether ABA moves radially outward ("piggy-backing" the redirection of water flux) after a xerobranching stimulus, we generated cross-sectional images of nlsABACUS2 root-tip tissues (Fig. 2) from three key zones. These zones included meristem and early elongation zone tissues (zone 1), mid- to late elongation and early differentiation zone tissues (zone 2), and fully differentiated root tissues (zone 3). Radial images of these three zones (taken at different time points during the xerobranching response) revealed contrasting patterns of ABA redistribution. For example, an increase in ABA levels can initially be detected in inner root tissues that is later seen in outer tissues in zone 2 (compare nlsABACUS2 signals in 3- versus 4-mm root-tip sections for zone 2 in Fig. 2), consistent with ABA movement from inner to outer root tissues. By contrast, radial cross sections through nlsABACUS2 zone 1 root tissues revealed an increase in ABA in outer (but not inner) tissues, which suggests that this signal originates from

[1]Plant and Crop Sciences, School of Biosciences, University of Nottingham, Nottingham, UK. [2]Earth and Life Institute, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium. [3]Sainsbury Laboratory, University of Cambridge, Cambridge, UK. [4]School of Plant Sciences and Food Security, Tel Aviv University, Tel Aviv, Israel. [5]Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium. [6]Center for Plant Systems Biology, VIB-UGent, 9052 Ghent, Belgium. [7]Crop Genetics, John Innes Centre, Norwich Research Park, Norwich, UK. [8]Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, SE-901 87 Umeå, Sweden. [9]Lancaster Environment Centre, Lancaster University, Lancaster, UK. [10]Division of Plant Science and Technology, University of Missouri, Columbia, MO, USA. [11]Department of Biosciences, University of Durham, Durham, UK.
***Corresponding author. Email: p.mehra@nottingham.ac.uk (P.M.); malcolm.bennett@nottingham.ac.uk (M.J.B.)**
†Present address: Oxford University Innovation, Oxford, UK.

phloem unloading in zone 2 and then undergoes redistribution through symplastic continuity. ABA accumulated much less in outer root tissues of zone 3 versus zones 1 and 2 after a xerobranching stimulus (Fig. 2). The results are consistent with the phloem unloading patterns of solutes described earlier (*19*), which suggests that zone 2 overlaps with the protophloem unloading zone, where solutes are transferred laterally through symplastic continuity. By contrast, zone 3 is active in solute translocation through phloem but inactive in phloem unloading (*19*). Another parallel study also reports similar expression patterns of nlsABACUS when exogenously applied ABA unloads through phloem (*18*). The pattern of elevated ABA in outer root tissues of zones 1 and 2 also coincides with the root-tip region, where new lateral root primordia initiate (termed the basal meristem or oscillatory zone) (*20*, *21*). Elevated levels of ABA in zones 1 and 2 rapidly decrease once root-tip tissues contact a new moisture source during the recovery phase (Fig. 2). Endogenous ABA levels in cotyledons of nlsABACUS2 plants did not change after a xerobranching stimulus, which suggests that ABA accumulation is locally confined to root tips that are exposed to air gaps (figs. S7 to S9). Hence, during a xerobranching response, ABA accumulates in the root zone that is associated with the earliest stage of lateral root development.

To test the function of the observed changes in ABA after a xerobranching stimulus, we examined the impact of mutations in key components of the ABA signaling machinery for this adaptive response. ABA-insensitive mutants disrupting ABA receptors (*pyr/pyl*) and ABA-activated SnRK2 kinases (*snrk2.2/2.3*, *snrk2.2/2.3/2.6*) exhibited lateral root branching in our xerobranching bioassay (Fig. 3A and figs. S10 and S11), whereas ABA-hypersensitive *pp2c* mutants failed to produce lateral roots in the air gap (fig. S12). Detailed analysis with *snrk* triple, septuple, and nonuple mutants as well as *snrk2.2* complementation lines revealed that *SnRK2.2* is required for xerobranching (Fig. 3D and fig. S11). Roots of a *snrk* nonuple mutant (which lacked 9 of 10 members of the SnRK2 family except SnRK2.2) behaved like WT in our xerobranching bioassay (fig. S11). Furthermore, a green fluorescent protein (GFP)–tagged SnRK2.2 reporter *pSnRK2.2:SnRK2.2-GFP* revealed that a xerobranching stimulus induced the fusion protein in the basal root meristem (Fig. 3, B and C, and fig. S13), which overlaps with the site of largest emission ratio changes of the ABA biosensor nlsABACUS2 (Fig. 1J). Hence, SnRK2.2 regulates ABA (and in turn, the xerobranching) response in *Arabidopsis* root tissues.

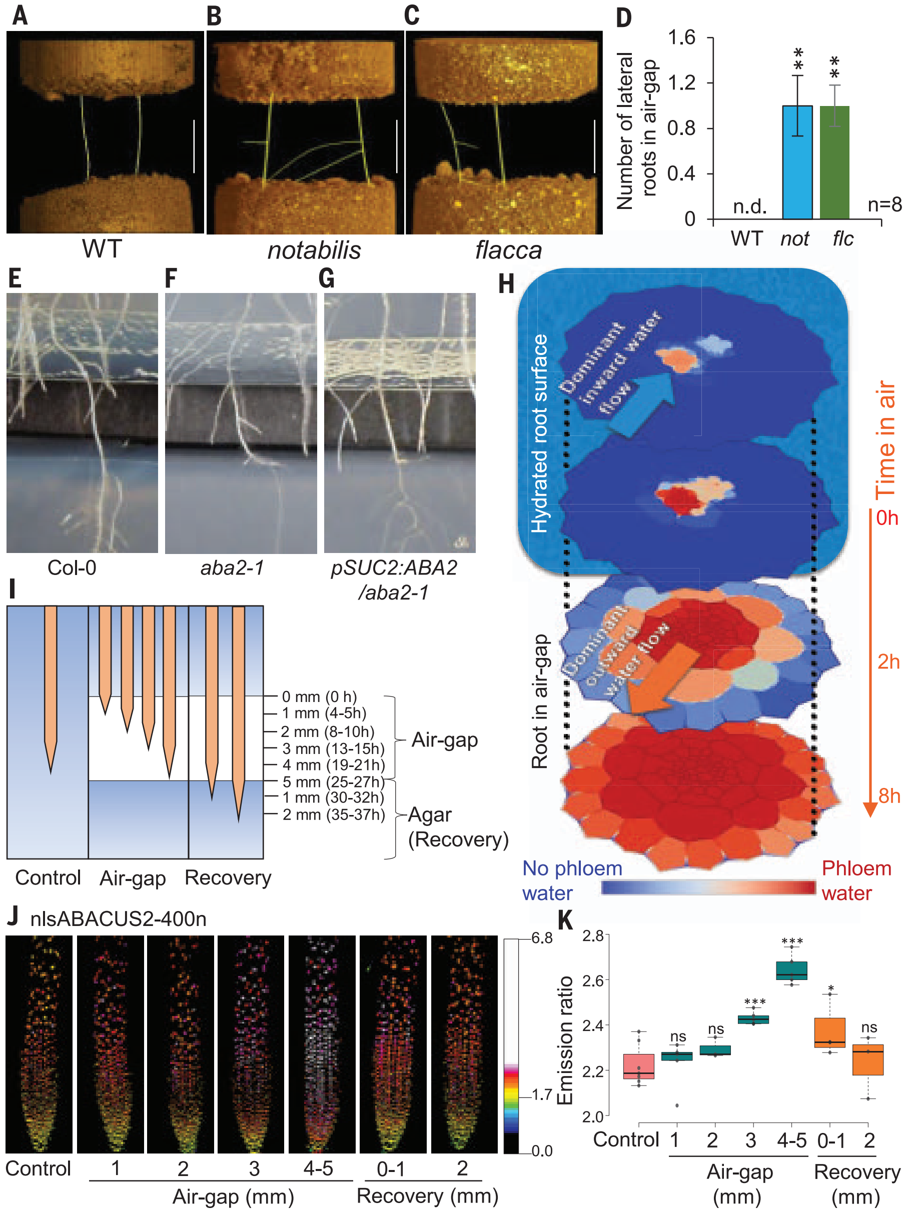


Fig. 1. ABA functions as a repressive signal during xerobranching response. (**A** to **D**) Representative CT images [(A) to (C)] and a bar graph (D) showing xerobranching defect in ABA-deficient tomato mutants (*notabilis* and *flacca*) versus WT (Ailsa Craig). Air gap, ~1.5 cm. Scale bar, 1 cm. n.d., not detected. (**E**) Agar-based air-gap assay system showing xerobranching response in *Arabidopsis* WT (Col-0) (air gap, ~5 mm). (**F**) *Arabidopsis* ABA biosynthetic mutant *aba2* produces lateral roots in air gap. (**G**) Expression of *ABA2* in phloem companion cells rescues xerobranching defect of *aba2*. (**H**) The MECHA hydraulic model predicts phloem as the main source of water in *Arabidopsis* root elongation zone only when located in an air gap. Simulated water advection maps show dominating phloem water in root epidermal cells within 8 hours. (**I**) For studying temporal response of reporters during xerobranching, primary root tips were grown to different lengths in air-gap (0 to 5 mm) and recovery (up to 2 mm in moist agar) conditions. (**J** and **K**) Emission ratio images (J) and box plots (K) of nlsABACUS2-400n reveal dynamic changes in endogenous ABA levels as root tips grow across air gap and after recovery. Color scale represents emission ratios. *$P \leq 0.05$; ***$P \leq 0.001$; ns, nonsignificant changes.

ABA regulates xerobranching by closing plasmodesmata

Genetic and biosensor-based studies revealed that ABA moves from its inner root (phloem) source to outer root tissues during a xerobranching response (Fig. 1, E to H and J; Fig. 2; and Fig. 3B). However, it remains unclear which root tissue(s) ABA targets to repress lateral root development during a xerobranching response. To determine this, we attempted to rescue the xerobranching defect of the ABA response mutant *snrk2.2/2.3* by expressing the WT *SnRK2.2* sequence under a series of root-tissue and/or zone-specific promoters (Fig. 3,

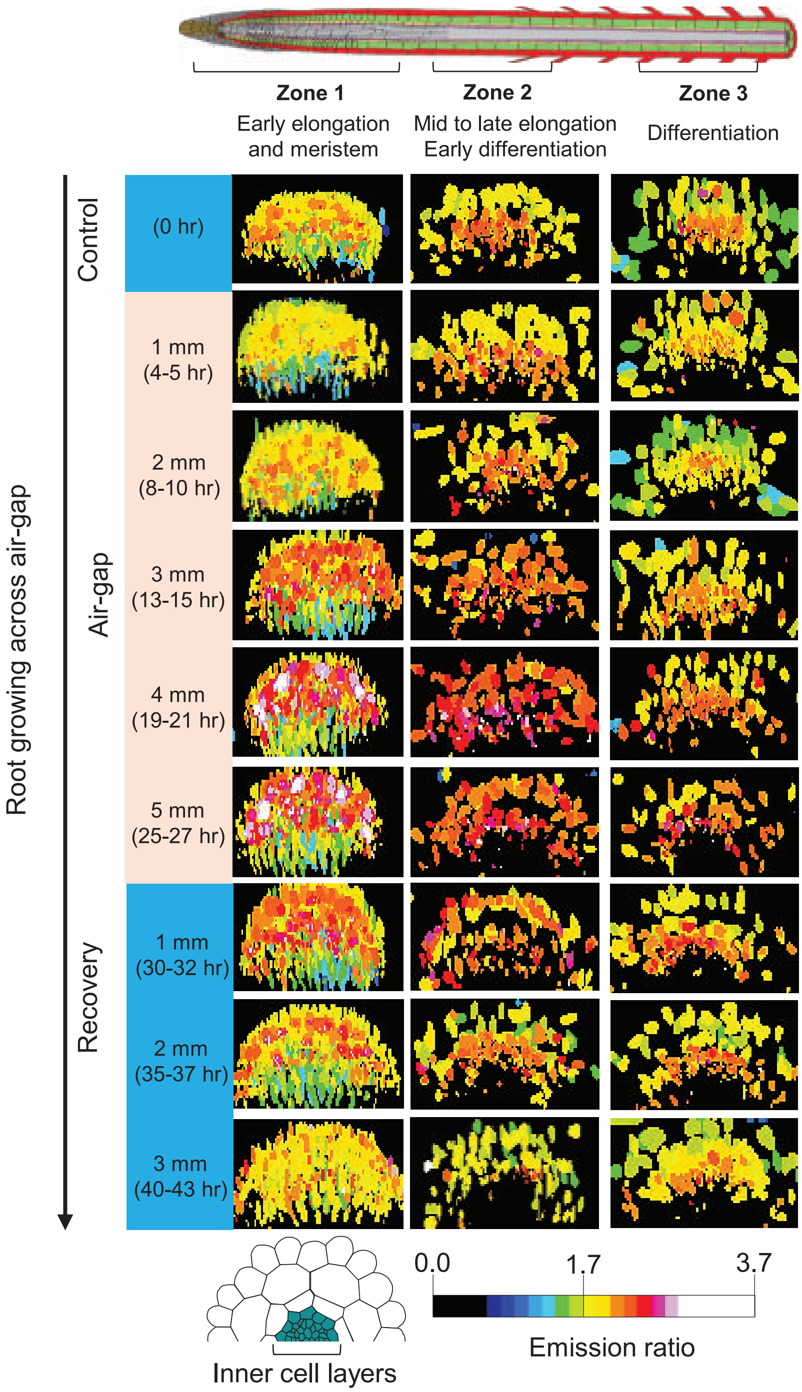


Fig. 2. ABA moves radially outward after a xerobranching stimulus. Ratio images of *Arabidopsis* roots expressing nlsABACUS2-400n reveal spatiotemporal distribution of ABA as the roots grow across an air gap. Each image is a representative maximum projection ratio map generated from a stack of several radial cross sections. Roots were imaged at different time points (indicated with hours) as they exit moist agar (control), grow in a 5-mm air gap, and reconnect with agar (recovery). Length (in millimeters) on the vertical axis indicates the length of root tips in air-gap and recovery conditions. For each time point, cross sections were generated from different zones of roots marked as zone 1 (early elongation and meristem), zone 2 (mid- to late elongation and early differentiation), and zone 3 (differentiation). Radial cross sections from zone 1 (compare cross sections from 1 to 5 mm) reveal an increase in ABA in outer (but not inner) tissues, which appears to originate from phloem unloading in zone 2 after a xerobranching stimulus. Color scale represents FRET emission ratios. Higher ratios correspond to higher ABA concentrations. Experiment was performed with at least four replicates per time point per zone (n = 4).

E to J). Expressing *SnRK2.2* in either root epidermal, cortical, endodermal, or basal meristem (but not lateral root cap) tissues rescued the *snrk2.2/2.3* xerobranching defect (Fig. 3, E to J). This result led us to ask, is there a common target important for xerobranching that the ABA signaling pathway regulates in each of these outer root tissues? ABA modulates hydraulic conductivity by inducing aquaporin expression to affect water movement out of the vasculature (*22*). To investigate possible roles of aquaporins in ABA-dependent repression of branching in air gaps, we analyzed a *pip* (plasma membrane intrinsic proteins) quadruple mutant in our xerobranching bioassay system. However, analysis of the *pip* mutant disrupting the most highly expressed *PIP* genes in roots did not show any significant difference compared with the WT control (fig. S14).

ABA has been reported to trigger reversible closure of intercellular pores, termed plasmodesmata (PD) (*23*). Does ABA regulate the xerobranching response by gating PD between root tissues? To address this, the symplastic fluorescent tracer carboxyfluorescein (CF) [derived from carboxyfluorescein diacetate (CFDA)] was used to monitor whether PD were open or shut in root tissues during a xerobranching response. Within plant cells, nonfluorescent CFDA is cleaved by cellular esterases to release membrane-impermeable fluorescent CF that travels exclusively through the PD. After CFDA application to cotyledons, we observed decreased protophloem unloading in *Arabidopsis* WT root tips that were exposed to a xerobranching stimulus (fig. S15). Delayed unloading was also observed in root tips of a *pSUC2:GFP* line that expresses free GFP in phloem companion cells (fig. S16). Collectively, both lines of evidence reveal that root tips experience decreased PD permeability after a xerobranching stimulus.

To assess the role of ABA in root PD gating during xerobranching, CFDA was locally applied to root tips. WT roots exhibited reduced intercellular fluxes of the tracer, which indicated closed PD after a xerobranching stimulus (Fig. 4, A and B). By contrast, CFDA treatment of root tips of the ABA response mutant *snrk2.2/2.3* revealed that PD remained open after a xerobranching stimulus (fig. S17). PD aperture is regulated by modifying the size of a callose ring encircling each end of the intercellular pore. The callose stain aniline blue revealed higher callose deposition in basal meristem cells of WT roots compared with *snrk2.2/2.3* after a xerobranching stimulus (fig. S18). Taken together, the evidence indicates that ABA regulates gating of PD in key root tissues during a xerobranching response.

To determine the function of PD gating during a xerobranching response, *Arabidopsis* mutants no longer able to gate their PD were characterized (figs. S19 and S20). PD aperture

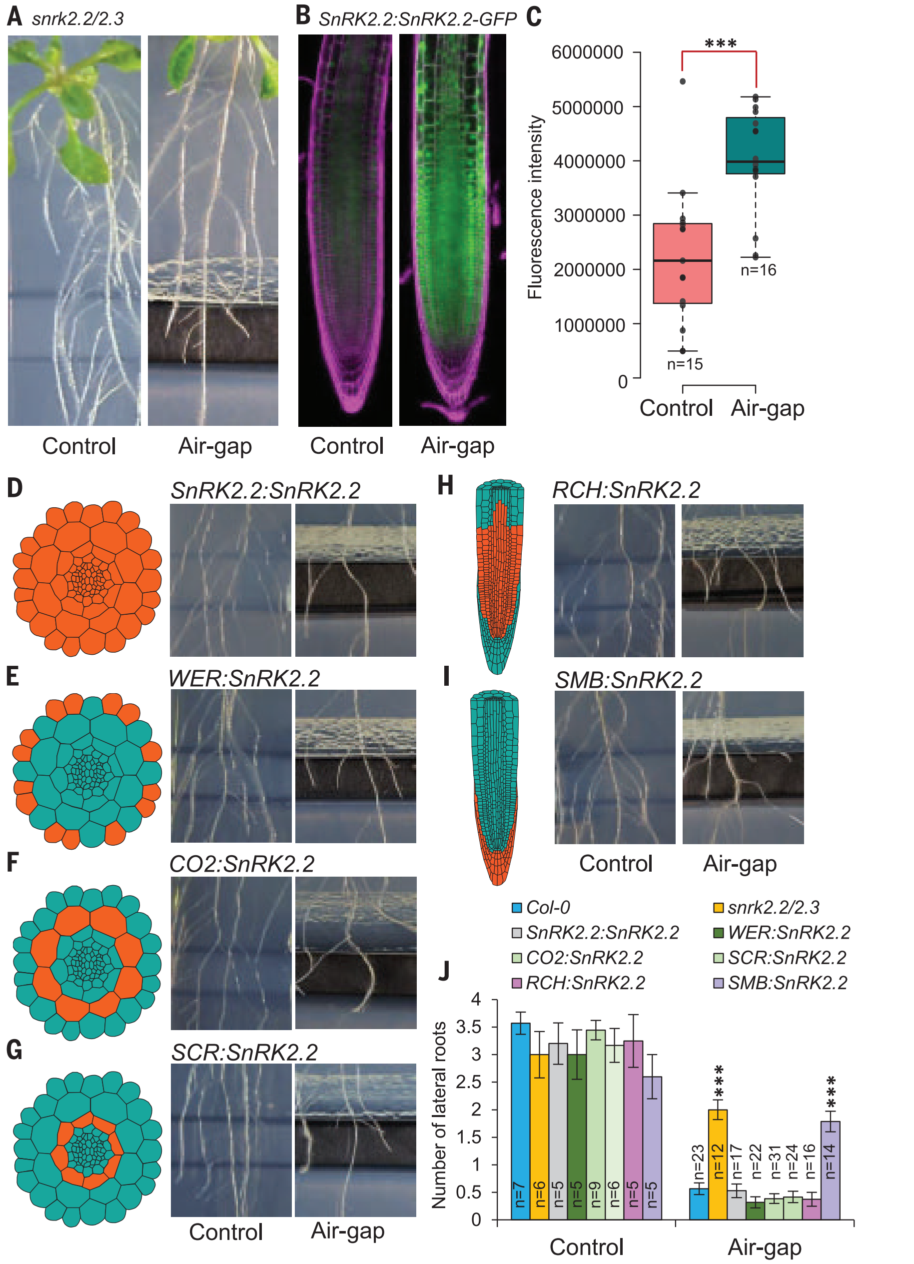


Fig. 3. Xerobranching is dependent on ABA perception in *Arabidopsis* primary root transition zone tissues. (**A**) ABA signaling mutant *snrk2.2/2.3* disrupts xerobranching in agar-based air-gap assay system. Air gap, ~5 mm. (**B** and **C**) Elevated response of *pSnRK2.2:SnRK2.2-GFP* in root tips subjected to air-gap versus control conditions. Scale bar, 100 μm. (**D**) Expression of *SnRK2.2* under its own promoter rescues the xerobranching defect of *snrk2.2/2.3*. (**E** to **G**) Functional complementation of *snrk2.2/2.3* by tissue-specific expression of *SnRK2.2* in epidermis, cortex, and endodermis using the *WER*, *CO2*, and *SCR* promoters, respectively. (**H**) Root zone–specific expression of *SnRK2.2* in root meristem and transition zone (using *RCH* promoter) rescues the xerobranching defect of *snrk2.2/2.3*. (**I**) Expression of *SnRK2.2* in columella and lateral root cap using SMB promoter in *snrk2.2/2.3* background. (**J**) Number of lateral roots produced by WT (Col-0), *snrk2.2/2.3*, and different tissue- and/or zone-specific rescue lines in the air-gap region versus the corresponding region of roots grown under control conditions. Orange color in root schematics defines the tissue and/or zone of expression of selected promoters used for complementation. Significant changes with respect to control (C) or WT (J) were calculated by Student's *t* test. ***$P \leq 0.001$.

is regulated through the action of enzymes, including a family of callose synthases and associated regulatory proteins such as PD-located proteins (PDLPs) (*24*). Mutating individual members of the *Arabidopsis* callose synthase gene family resulted in nine of these lines exhibiting a xerobranching defect, including *CALS7*, which is primarily associated with phloem sieve pores (fig. S19) (*25*), as did mutating both *PDLP2* and *PDLP3* genes (fig. S20). Hence, blocking the ability to close PD disrupts a root's ability to activate a xerobranching response.

How does ABA trigger PD closure after a xerobranching stimulus? Transgenic lines expressing GFP reporters that were fused to either PDLP2 or PDLP3 proteins under their native promoters were up-regulated in response to a xerobranching stimulus (Fig. 4, C and D, and figs. S21 to S23). In the case of PDLP3, its mRNA and *PDLP3:PDLP3-GFP* reporter were up-regulated in root tip and, specifically, basal meristem cells during a xerobranching response, exhibiting a punctate pattern in apical, basal, and side walls, consistent with its PD regulatory function (Fig. 4, C and D, and figs. S21 and S23). Treating roots of *PDLP3:PDLP3-GFP* with low levels of ABA (50 nM) phenocopied a xerobranching stimulus (fig. S24), in agreement with the presence of several ABA-responsive elements (ABREs) in its promoter sequence (table S1). We also noted that many other *PDLP* and *callose synthase* genes contained ABREs in their promoter sequences (tables S1 and S2). Comparing spatiotemporal response curves of nlsABACUS2 and PDLP3-GFP (during xerobranching and subsequent recovery) reveals that ABA accumulation precedes PDLP3 induction (Fig. 1K and fig. S23). To test the functional importance of ABA-regulated PDLP up-regulation during xerobranching, we characterized the ABA-insensitive mutant *abi1-1*. The dominant negative *abi1-1* allele disrupts ABA-dependent PD closure and maintains high frequencies of open PD (*26*). Unlike WT, *abi1-1* roots exhibit branching in air gaps (fig. S25). However, ectopic expression of *PDLP1* (using *35S* rather than its native promoter) rescued the xerobranching defect of *abi1-1* (fig. S25). Hence, ABA triggers PD closure after a xerobranching stimulus by up-regulating expression of *PDLP*s. Additionally, GAL4-mediated transactivation of *abi1-1* in root ground tissues (cortex and endodermis) led to maximal disruption of xerobranching response as compared with induction of *abi1-1* in single root cell layers (fig. S26). Hence, ABA-mediated PD closure in root ground tissues is essential for a xerobranching response.

PD closure blocks inward auxin movement

How does ABA-mediated PD gating disrupt lateral root branching? Our reporter studies reveal that closing PD during a xerobranching response blocks the inward (symplastic) movement of another plant hormone signal, auxin, which is required to initiate lateral root branching. Exposing roots of the *DR5:VENUS* auxin reporter line to a xerobranching stimulus results in an elevated auxin response in epidermal cells within the root basal meristem zone (Fig. 4, E and F). The altered *DR5:VENUS* reporter pattern is consistent with xerobranching-induced PD closure in basal meristem epidermal cells transiently blocking the radially inward symplastic movement of auxin (Fig. 4, E and F, and fig. S27).

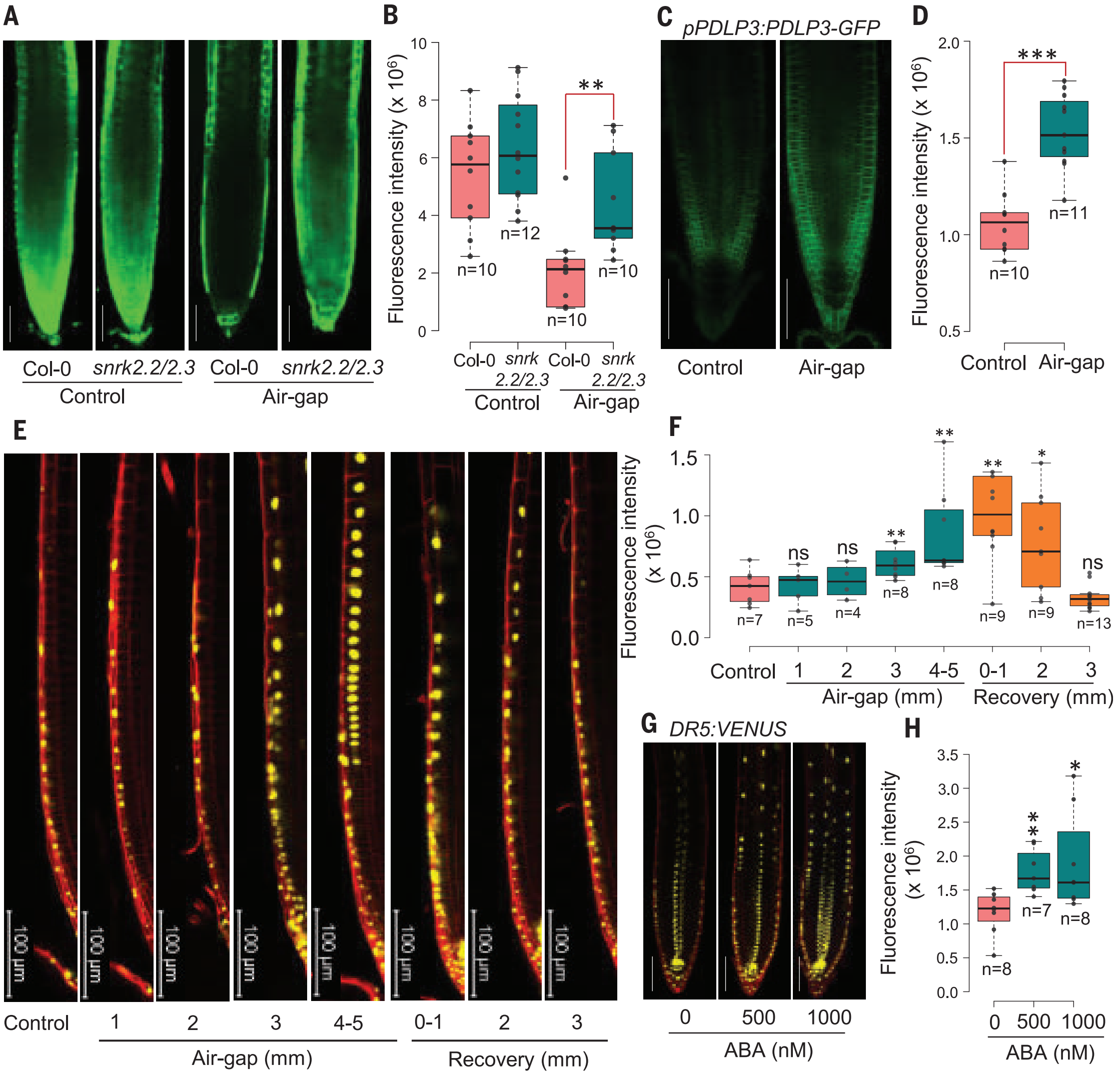


Fig. 4. ABA blocks radial auxin flow by reducing PD permeability causing xerobranching. (**A**) Slower migration of symplastic tracer CF in WT *Arabidopsis* (Col-0) primary root tips subjected to xerobranching stimulus versus control conditions. CF migration was indistinguishable in ABA signaling mutant *snrk2.2/2.3* with or without the xerobranching stimulus. Root tips were treated with CFDA before imaging. (**B**) CF fluorescence intensity in WT and *snrk2.2/2.3* root tips exposed to air-gap versus control conditions. (**C**) Xerobranching stimulus induces expression of *pPDLP3:PDLP3-GFP*. (**D**) Box plots showing significantly enhanced response of *pPDLP3:PDLP3-GFP* in air-gap versus control conditions. (**E** and **F**) Auxin response reporter *DR5:VENUS* shows higher fluorescence signal in epidermis of root tips exposed to air-gap versus control and recovery conditions. (**G** and **H**) ABA treatments phenocopy *DR5:VENUS* response during xerobranching. Scale bar, 100 μm. *$P \leq 0.05$; **$P \leq 0.01$; ***$P \leq 0.001$; ns, nonsignificant (Student's *t* test).

The auxin reporter DII-VENUS also revealed higher auxin abundance in root tips during exposure to a xerobranching stimulus (figs. S27 and S28). The effect of a xerobranching stimulus on *DR5:VENUS* and DII-VENUS can be mimicked by exogenous application of ABA to roots (Fig. 4, G and H, and fig. S29). ABA-mediated PD gating during a xerobranching response blocks radial inward auxin movement and lateral root induction. Consistent with this model, switching off ABA signaling in the *abi1-1* mutant (which ensured that PD remained constitutively open) blocked any change in DII-VENUS auxin response after a xerobranching stimulus (fig. S30). Suppression of lateral root initiation in roots exposed to a xerobranching stimulus was visualized by using the *DR5:LUC* reporter, which normally demarks groups of lateral root stem cells in the basal meristem (or oscillation zone) that go on to form new branches in control roots, but these *DR5:LUC* oscillations were blocked in roots exposed to a xerobranching stimulus (fig. S31). Hence, the xerobranching response blocks radially inward symplastic auxin movement that is required for lateral root induction.

Root auxin distribution was originally thought to be exclusively determined by the combined activities of AUX1/LAX, ABCB, and PIN auxin influx and efflux carriers (*27*). To determine the impact of xerobranching on carrier-mediated auxin transport, reporter lines for key auxin influx (*pAUX1:AUX1-YFP*) or efflux (*pPIN2:PIN2-GFP*) carriers expressed in the root elongation zone were analyzed (fig. S32). No spatial changes were detected in AUX1-YFP or PIN2-GFP expression in the outermost root tissues after a xerobranching stimulus, consistent with the observed elevated auxin response (Fig. 4, E and F, and fig. S27) being the result of gating of hormone movement through PD rather than through modulation of apoplastic auxin transport (fig. S32). Ectopic expression of either *AUX1* or *PIN2* in every elongation zone tissue, to create a synthetic radial apoplastic auxin pathway, successfully bypassed the PD-mediated restriction of symplastic auxin flow and led to lateral root branching in air gaps (fig. S33).

Collectively, multiple lines of evidence establish a role for ABA-mediated PD gating of radial symplastic auxin flow to regulate xerobranching. Roots appear to sense water availability by comobilizing water and hormones through PD. If water, ABA, and auxin moved through substrate-specific carriers (aquaporins, ABCG, and PINs), their comobilization would be broken. Instead, because PD are large intercellular pores that are nonselective, radial hydraulic fluxes comobilize hormones either inward (such as auxin) or outward (such as ABA) when water is available or not, respectively. Thus, auxin and ABA can be considered as hydrosignals when comobilized through PD with water.

Our study has revealed that xerobranching is a widely conserved, ABA-dependent root adaptive response to localized loss of contact with soil water in eudicot and monocot species. We also report the molecular and cellular mechanisms that enable plant roots to block lateral root formation after experiencing a xerobranching stimulus (see schematic in Fig. 5). Under optimal moisture conditions,

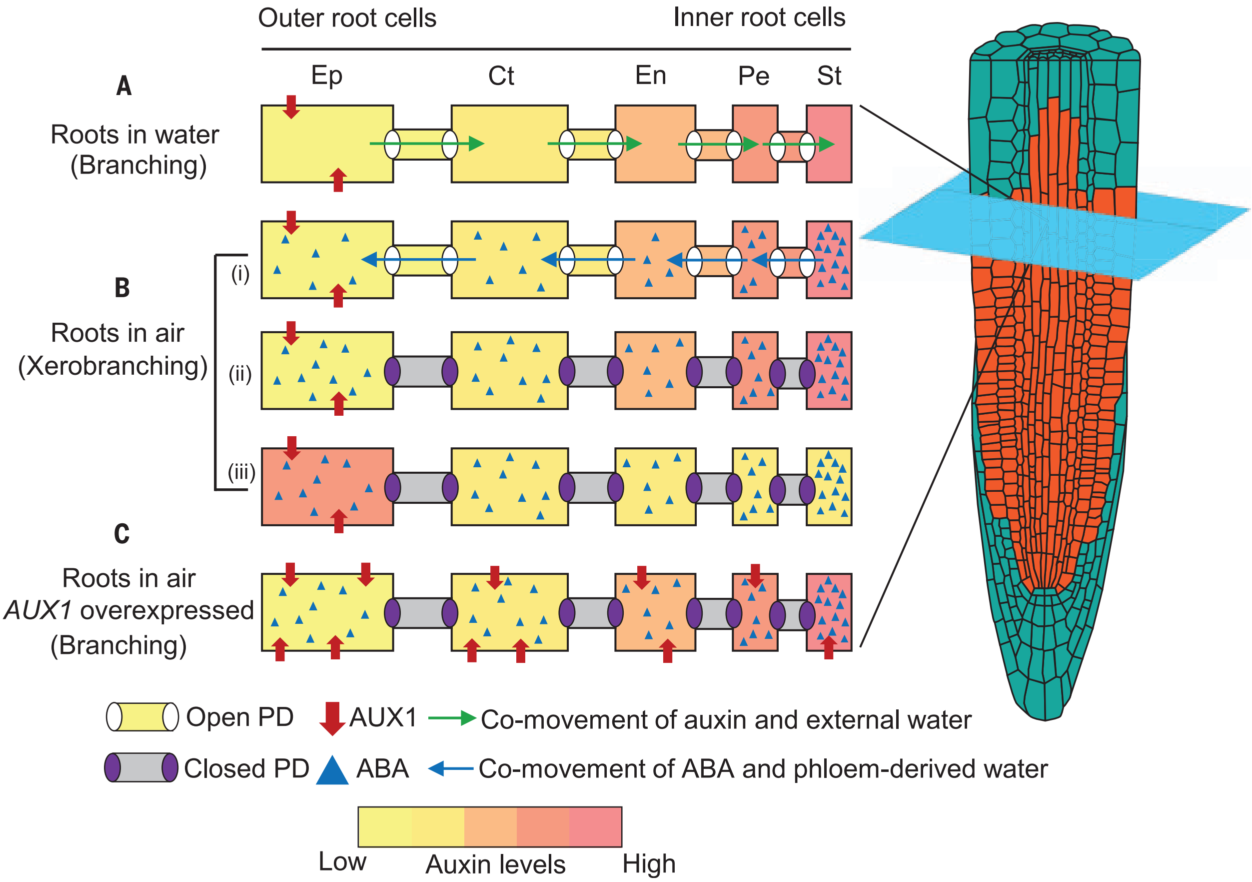


Fig. 5. Dynamic hormone redistribution determines root branching or xerobranching response to external water availability. Schematics illustrate cellular auxin, ABA, and water fluxes in root basal meristem (highlighted cell files in transverse root section) exposed to water or air. (**A**) Under normal moisture conditions, roots cotransport water and auxin radially inward through PD, which delivers the key hormone signal (auxin) to pericycle cells, where lateral root initiation is triggered. (**B**) Xerobranching stimulus triggers rise in ABA levels in inner root tissues (stele), which suppresses root branching. In the transient absence of external water, phloem-derived water comobilizes ABA from inner (stele) to outer root cells (epidermis) through PD (i). ABA induces closure of PD through *PDLP*s and *CALs* (ii) disrupting symplastic radial movement of auxin to pericycle (iii). (**C**) Ectopic expression of auxin carriers such as AUX1 in all elongation zone tissue creates a synthetic radial apoplastic auxin pathway, which bypasses the disrupted symplastic auxin flow and thus causes root branching in air gap. Ct, cortex; En, endodermis; Ep, epidermis; Pe, pericycle; St, stele.

roots cotransport auxin and water inward between root basal meristem tissues through PD and thus provide a symplastic pathway to radially mobilize this key hormone signal with hydraulic fluxes from epidermal to pericycle cells, where auxin triggers lateral root initiation. However, when the external source of water is transiently unavailable, root tips rely on an internal (phloem) source of water to continue growing (*15*) and couple its outward radial flow with movement of the hormone ABA. This key water stress signal reduces PD aperture, disrupting radial movement of the branching signal auxin to lateral root "stem cells" in the pericycle. Restoring an external water source (during the recovery phase) rapidly reduces ABA levels in the outermost root tissues (>3 hours) and thus restores open PD and symplastic auxin transport. Once reconnected with an external water source, the abundance and distribution of these signals reset in root-tip tissues to pre–xerobranching-stimulus levels. These dynamic regulatory events—coupling changes in hydraulic fluxes with hormone redistribution, which we term hydrosignaling—enable plant roots to calibrate spatial root branching responses in heterogeneous soil environments, which are the norm rather than the exception. Climate change is likely to exacerbate water scarcity (*9*, *10*); thus, that water in soils will be even more heterogeneously distributed in the future than at present is highly likely. Insight into the molecular and cellular mechanisms through which roots acclimate under these conditions may improve the climate resilience of future crops.

REFERENCES AND NOTES

1. B. Muller *et al.*, *Trends Plant Sci.* **24**, 810–825 (2019).
2. J. P. Lynch, *Plant Physiol.* **156**, 1041–1049 (2011).
3. H. Motte, T. Beeckman, *J. Exp. Bot.* **70**, 785–793 (2019).
4. E. C. Morris *et al.*, *Curr. Biol.* **27**, R919–R930 (2017).
5. A. Zhan, H. Schneider, J. P. Lynch, *Plant Physiol.* **168**, 1603–1615 (2015).
6. R. Karlova, D. Boer, S. Hayes, C. Testerink, *Plant Physiol.* **187**, 1057–1070 (2021).
7. B. Orman-Ligeza *et al.*, *Curr. Biol.* **28**, 3165–3173.e5 (2018).
8. Y. Bao *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 9319–9324 (2014).
9. D. B. Lobell, S. M. Gourdji, *Plant Physiol.* **160**, 1686–1697 (2012).
10. Y. Grusson, I. Wesström, E. Svedberg, A. Joel, *Agric. Water Manage.* **249**, 106766 (2021).
11. A. J. Thompson *et al.*, *Plant Cell Environ.* **27**, 459–471 (2004).
12. I. B. Taylor, R. S. T. Linforth, R. J. Al-Naieb, W. R. Bowman, B. A. Marples, *Plant Cell Environ.* **11**, 739–745 (1988).
13. B. C. Tan, S. H. Schwartz, J. A. D. Zeevaart, D. R. McCarty, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12235–12240 (1997).
14. T. Kuromori, E. Sugimoto, K. Shinozaki, *Plant Physiol.* **164**, 1587–1592 (2014).
15. B. S. Wiegers, A. Y. Cheer, W. K. Silk, *Plant Physiol.* **150**, 2092–2103 (2009).
16. V. Couvreur *et al.*, *Plant Physiol.* **178**, 1689–1703 (2018).
17. F. C. Pascut *et al.*, *Nat. Commun.* **12**, 4682 (2021).
18. J. Rowe *et al.*, bioRxiv 2022.10.19.512731 [Preprint] (2022).
19. T. J. Ross-Elliott *et al.*, *elife* **6**, e24125 (2017).
20. I. De Smet *et al.*, *Development* **134**, 681–690 (2007).
21. M. A. Moreno-Risueno *et al.*, *Science* **329**, 1306–1311 (2010).
22. K. Prado *et al.*, *Plant Cell* **25**, 1029–1039 (2013).
23. Y. Benitez-Alfonso, *Plant Cell Physiol.* **60**, 713–714 (2019).
24. C. L. Thomas, E. M. Bayer, C. Ritzenthaler, L. Fernandez-Calvino, A. J. Maule, *PLOS Biol.* **6**, e7 (2008).
25. B. Xie, X. Wang, M. Zhu, Z. Zhang, Z. Hong, *Plant J.* **65**, 1–14 (2011).

26. S. Tylewicz *et al*., *Science* **360**, 212–215 (2018).
27. J. Petrášek, J. Friml, *Development* **136**, 2675–2688 (2009).

ACKNOWLEDGMENTS

We thank P. L. Rodriguez (CSIC-UPV, Spain) for sharing *pyr/pyl1124589*, *pyr/pyl1124578*, and *pyr/pyl112458* mutants; H. Fujii (University of Turku, Finland) for sharing *snrk2.2/3/6*, *snrk2.1/4/5/7/8/9/10*, and *snrk2.1/3/4/5/6/7/8/9/10* mutants; and R. Miao (Fujian Agriculture and Forest University, China) for sharing *hab1-1abi1-2pp2ca*, *hab1-1abi1-2abi2-2*, and *Qabi2-2* mutants. We thank L. Kalmbach, Y. Helariutta (SLCU, UK), and N. Geldner (University of Lausanne, Switzerland) for sharing *pSUC2:GFP* seeds. We also thank Z. Gong and Y. Wang (China Agricultural University) for kindly sharing *abi1-1* (Col-0) seeds. We acknowledge C. Luschnig (University of Natural Resources and Life Sciences – BOKU, Vienna) and J. Friml (IST, Austria) for sharing *eir1-4/PIN2:PIN2-GFP* and *eir1-4/35S:PIN2-GFP* lines. We thank F. Augstein and A. Carlsbecker (Uppsala University, Sweden) for providing *UAS:abi1-1* and driver lines. We thank F. Chaumont (Louvain-la-Neuve, Belgium) and A. Schäffner (Helmholtz Zentrum München, Germany) for sharing the aquaporin mutant. **Funding:** This study was supported by a European Molecular Biology Organization Long-Term Fellowship (ALTF 1140-2019 to P.M.); European Union's Horizon 2020 research and innovation program, Marie Skłodowska-Curie, XEROBRANCHING (891262 to P.M.); Biotechnology and Biological Sciences Research Council Discovery Fellowship (BB/V00557X/1 to B.K.P.); the Biotechnology and Biological Sciences Research Council (BB/T001437/1 to M.J.B. and A.S.; BB/V003534/1 to M.J.B., A.S., and J.B.; BB/W008874/1 to M.J.B. and C.J.S.; BB/W015080/1 to M.J.B.; and BB/P018572/1 to A.M.J.); European Research Council grant FUTUREROOTS (294729 to M.J.B.); European Research Council, SUMOrice (grant no. 310235 to A.S.); Fonds de la Recherche Scientifique – FNRS (to V.C.); the Swedish Research Council (no. 2020-03522 to R.P.B.); European Research Council, RobustHormoneTrans (757683 to E.S.); the NSF Plant Genome Research Program (IOS-1444448 to R.E.S.); the Gatsby Charitable Foundation (to A.M.J. and J.R.); European Research Council grant INTERCELLAR (725459 to C.F.); and the Biotechnology and Biological Research Council Institute Strategic Programme, Plant Health (BBS/E/J/000PR9796 to C.F.). **Author contributions:** Conceptualization: P.M. and M.J.B. Methodology: P.M., B.K.P., V.C., J.R., and H.D.G. Investigation: P.M., B.K.P., D.M., J.B., N.L., V.C., J.R., M.A., H.D.G., E.M., C.J.S., and R.S. Funding acquisition: P.M. and M.J.B. Supervision: M.J.B. Writing – original draft: M.J.B. and P.M. Writing – review & editing: P.M., B.K.P., D.M., J.B., N.L., V.C., J.R., M.A., H.D.G., E.M., C.J.S., R.S., S.J.M., C.F., T.B., R.P.B., E.S., A.M.J., I.C.D., R.E.S., A.S., X.D., and M.J.B. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** No restrictions are placed on materials, such as materials transfer agreements. All data are available in the main text or the supplementary materials. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add3771
Materials and Methods
Figs. S1 to S33
Tables S1 and S2
References (*28–59*)
MDAR Reproducibility Checklist

Submitted 8 June 2022; accepted 24 October 2022
10.1126/science.add3771

REPORTS

3D PRINTING

Mechanical nanolattices printed using nanocluster-based photoresists

Qi Li[1]*†, John Kulikowski[1]†, David Doan[1]†, Ottman A. Tertuliano[1], Charles J. Zeman IV[2], Melody M. Wang[1], George C. Schatz[2], X. Wendy Gu[1]*

Natural materials exhibit emergent mechanical properties as a result of their nanoarchitected, nanocomposite structures with optimized hierarchy, anisotropy, and nanoporosity. Fabrication of such complex systems is currently challenging because high-quality three-dimensional (3D) nanoprinting is mostly limited to simple, homogeneous materials. We report a strategy for the rapid nanoprinting of complex structural nanocomposites using metal nanoclusters. These ultrasmall, quantum-confined nanoclusters function as highly sensitive two-photon activators and simultaneously serve as precursors for mechanical reinforcements and nanoscale porogens. Nanocomposites with complex 3D architectures are printed, as well as structures with tunable, hierarchical, and anisotropic nanoporosity. Nanocluster-polymer nanolattices exhibit high specific strength, energy absorption, deformability, and recoverability. This framework provides a generalizable, versatile approach for the use of photoactive nanomaterials in additive manufacturing of complex systems with emergent mechanical properties.

Advances in three-dimensional (3D) printing rely on the invention of new printing techniques (*1*, *2*) as well as the creation of compatible feedstock materials such as resins, filaments, powders, and photoresists (*3*, *4*). A current frontier is the printing of complex materials with arbitrary 3D nanoarchitectures. These structures are of particular interest as next-generation mechanical metamaterials (*5*) that combine nanoscale material size effects with lightweight lattice structures, which leads to optimized strength (*6*, *7*), stiffness (*8*), or energy absorption per weight (*9*). Improved performance can be achieved by printing advanced materials with complex internal nanofeatures, such as nanocomposites with optimized spacing and interactions between soft and hard components, or the hierarchical graded porosity and structural anisotropy of natural structural materials (*10*). Fabrication of these complex nanostructures is currently limited by existing printing techniques and lack of appropriate feedstock.

Two-photon (or multiphoton) lithography can achieve polymeric nanostructures with almost arbitrary geometry (*1*, *2*). Two-photon resists generally use photoinitiators that absorb photons to form reactive species (e.g., radicals) that initiate polymerization. High-performance two-photon initiators can greatly enhance printing speed and resolution (*11*, *12*). Currently, almost all two-photon initiators are organic molecules that do not enhance the properties and structural complexity of the final printed products (*13*, *14*). In addition, each organic photoinitiator is typically efficient only for a single type of monomer. Composite nanostructures can be printed by adding metal ions or inorganic particles to existing two-photon resists. This strategy is robust for simple nanostructures (e.g., nanowires, 2D patterns) (*15*) but ineffective for complex mechanical nanolattices, as performance suffers when small structural flaws are present (*5*, *16*). Uncontrollable metal ion reduction and particle growth lead to structural flaws and reduced feature quality, and aggregated particles interfere with light propagation (*5*, *16*). Composite nanolattices can also be fabricated by depositing an inorganic layer onto a printed polymeric scaffold (*17*, *18*). This approach results in impressive mechanical performance but requires a time-consuming, multistep fabrication process and is limited to core-shell structures.

In the past decade, ultrasmall metal nanoclusters 1 to 2 nm in size have been synthesized with precise atomic structures (*19*). These quantum-confined nanoclusters show nonmetallic, molecule-like electronic structures and optical properties. Metal nanoclusters consist of metal atoms with zero partial charge, which avoids the high energy requirements typical of metal reduction. The ultrasmall size of the nanoclusters circumvents the light scattering that would occur with larger metal particles. The densities of many metal nanoclusters are much lower than their bulk metal counterparts (*19*). Therefore, ultrasmall nanoclusters are prime candidates for replacement of traditional organic molecules as photoinitiators and substitutes for metal ions or particles as inorganic precursors. By doing so, fast printing of nanocomposite nanolattices with high

[1]Department of Mechanical Engineering, Stanford University, Stanford, CA 94305, USA. [2]Department of Chemistry, Northwestern University, Evanston, IL 60208, USA.
*Corresponding author. Email: qilistan@stanford.edu (Q.L.); xwgu@stanford.edu (X.W.G)
†These authors contributed equally to this work.

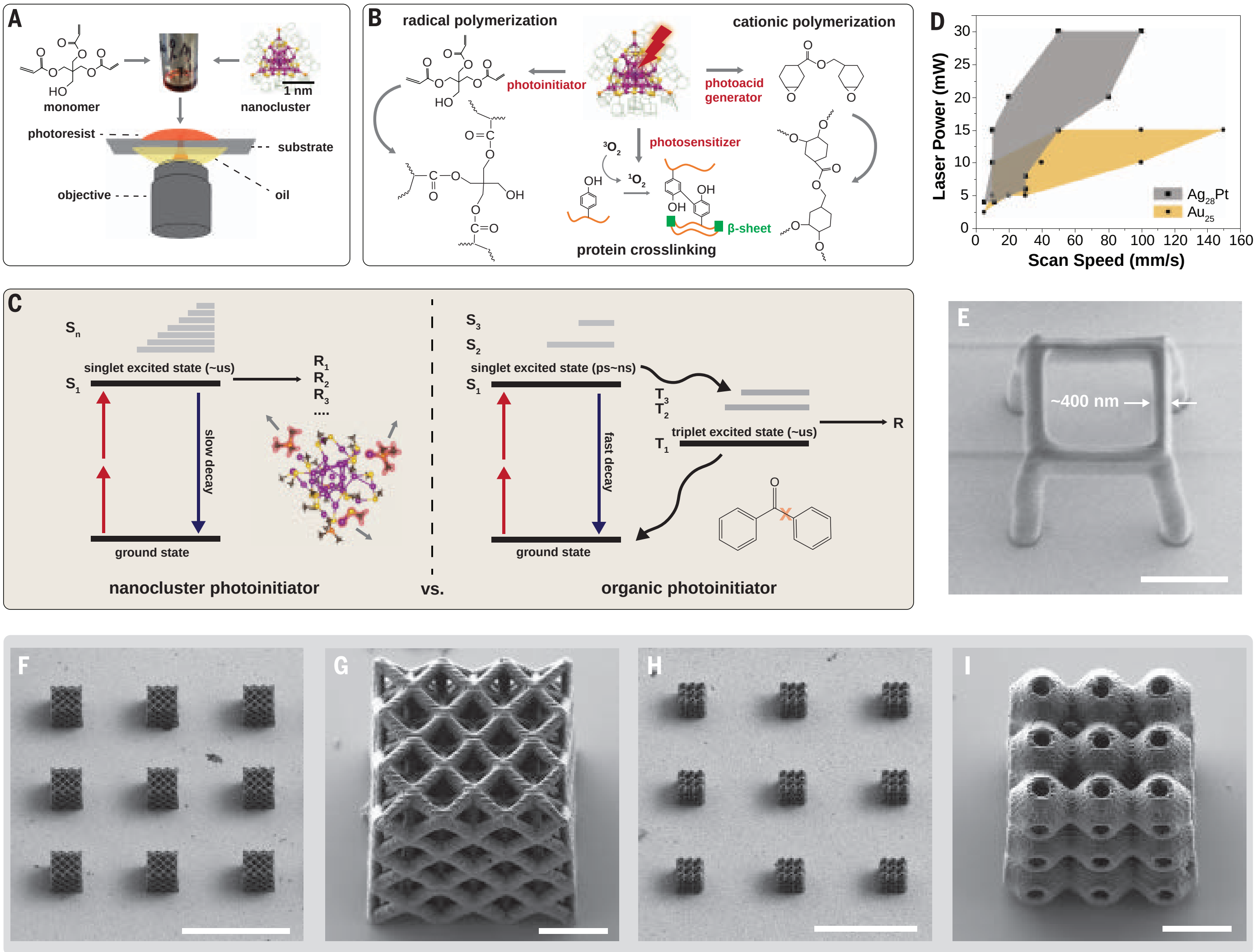


Fig. 1. Photochemistry and printability of nanocluster-based photoresist. (**A**) Photoresist composition and printing setup. (**B**) Schematic of photopolymerization of different classes of monomers during two-photon lithography. The atomic structure of the nanocluster ($Ag_{28}Pt$) is reproduced from (*20*). (**C**) Comparison of photochemistry of nanocluster photoinitiators (left) with typical organic photoinitiators (right). The left inset shows a snapshot from an excited-state Born-Oppenheimer molecular dynamics simulation of an $Ag_{28}Pt$ nanocluster, showing diverse fragmentation pathways (orange shading and gray arrows). The right inset shows a typical organic photoinitiator and its bond-dissociation position (orange X). (**D**) Fabrication window (laser power and scanning speed) for $Ag_{28}Pt$ and Au_{25} photoresists. (**E**) Open-faced table fabricated using $Ag_{28}Pt$ photoresist. (**F** and **G**) Octet lattices and (**H** and **I**) Schwarz primitive (SP) lattices made with Au_{25} photoresist. Scale bars: (E) 2 μm, (F) 100 μm, (G) 10 μm, (H) 100 μm, (I) 10 μm.

structural quality, fidelity, and complexity can be achieved.

Our nanocluster-based photoresist consists only of metal nanoclusters, monomers [e.g., an acrylate monomer such as pentaerythritol triacrylate (PETA) or epoxy monomer such as 3,4-epoxycyclohexylmethyl 3,4-epoxycyclohexanecarboxylate (EEC)] (Fig. 1A), and a solvent (2-propanol), which increases the solubility of nanoclusters in the photoresist. We find that nanoclusters can serve as photoinitiators for radical polymerization, photoacid generators for cationic polymerization, and photosensitizers that promote singlet oxygen formation to induce the crosslinking of proteins (Fig. 1B). The $Ag_{28}Pt$ nanocluster (*20*) and the rod-shaped Au_{25} nanocluster (*21*) were selected as the two-photon initiators in this study. The two-photon absorption cross sections of $Ag_{28}Pt$ and Au_{25} were found to be >30 Goeppert-Mayer units (GM) and >100 GM at sub–800-nm wavelengths, respectively, as determined by quadratic response time-dependent density functional theory calculations (fig. S1). Furthermore, these two nanoclusters have stable, long-living S_1 excited states (lifetime ~3 μs) (*21*), which are beneficial for the generation of radical or other reactive species (Fig. 1C, left). This is different from organic two-photon initiators in which the S_1 state is short lived (ps-ns). In organic two-photon initiators, intersystem crossing to triplet states is generally necessary for the generation of reactive species (*13*, *14*), which may reduce overall photoinitiation efficiency (Fig. 1C, right). More importantly, compared with molecular photoinitiators that usually have limited fragmentation pathways, nanoclusters provide for a greater variety of bond scissions (Fig. 1C, left, and movies S1 and S2), which lead to increased reactivity with different reagents.

The printability of nanocluster-acrylate photoresists was systematically evaluated. Squares and lines were fabricated as test structures to investigate the required scanning speed, threshold laser power, and smallest achievable feature size of the nanocluster-acrylate photoresist. For 5 wt % $Ag_{28}Pt$ photoresist

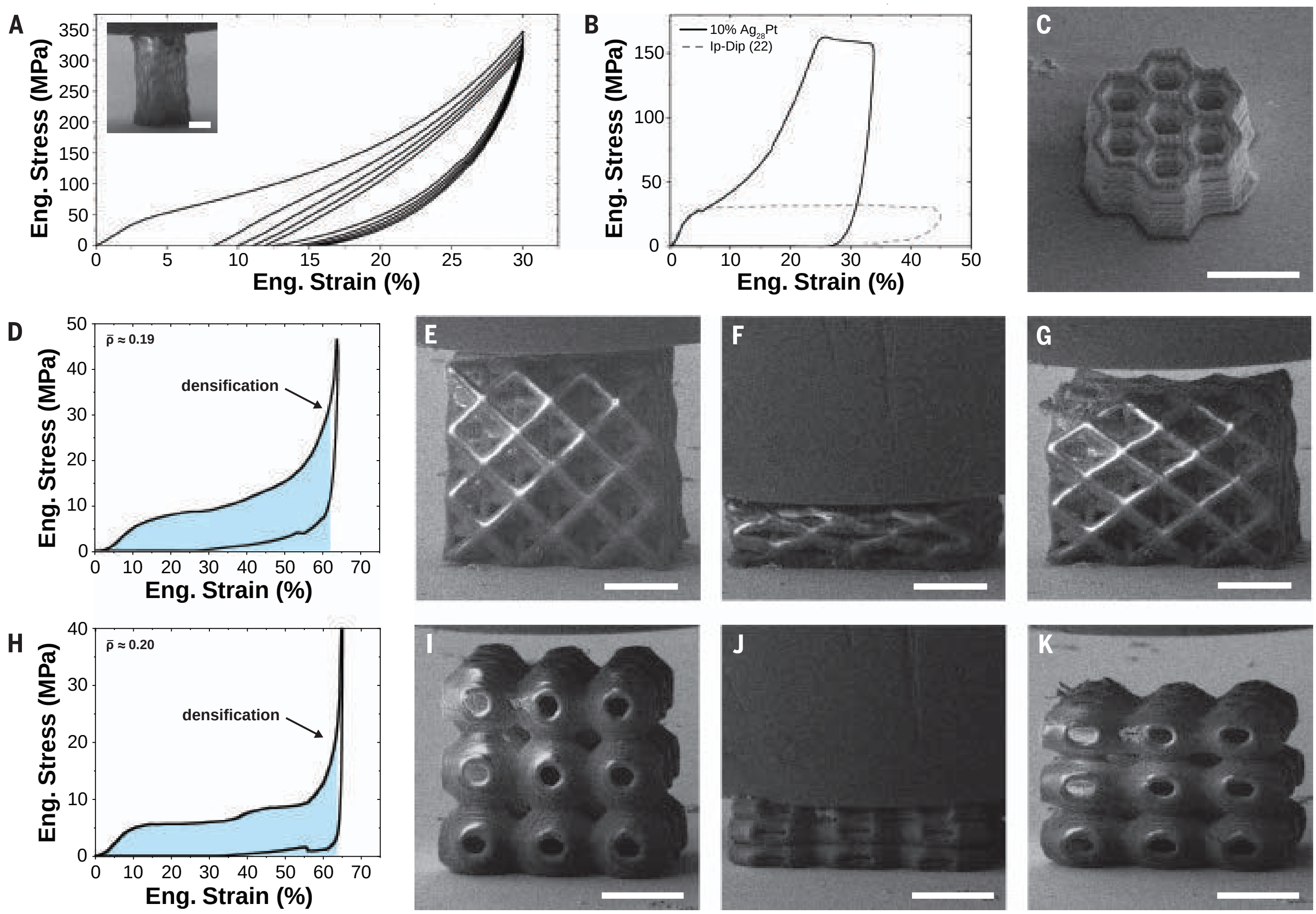


Fig. 2. Mechanical behavior of nanocluster-polymer nanolattices. (**A**) Engineering stress-strain curve of pillars under cyclic loading. Inset shows SEM image of a pillar before compression. (**B**) Engineering stress–engineering strain curves and (**C**) SEM image of honeycomb lattices. Stress-strain curve for Ip-Dip honeycomb is from (*22*). (**D**) Engineering stress–engineering strain curve for an octet lattice. (**E** to **G**) In situ SEM compression of an octet lattice. (**H**) Engineering stress–engineering strain curve for SP lattice. (**I** to **K**) In situ SEM compression of an SP lattice. Scale bars: (A) 2 μm, (C) 8 μm, (E to G and I to K) 10 μm.

(5 mg $Ag_{28}Pt$ nanoclusters and 95 mg PETA), square structures can be fabricated at laser power as low as 4 mW and scanning speeds up to 100 mm/s (Fig. 1D and figs. S2 and S3). The fabricated structure shows photoluminescence with a maximum peak at 720 nm (fig. S4), similar to $Ag_{28}Pt$ nanoclusters in solution, which suggests that the nanoclusters remain intact in the printed structure. Lines were fabricated with a lateral resolution of ~200 to 250 nm (fig. S5). Similar performance was shown by 8 wt % Au_{25} photoresists, with threshold power down to 2.5 mW and scanning speed up to 150 mm/s (Fig. 1D). Open-faced tables, octet lattices, and Schwarz primitive (SP) lattices made using $Ag_{28}Pt$ and Au_{25} photoresists are shown in Fig. 1, E to I. Figure 1E shows that freestanding 3D features can be as small as 400 nm. The minimal strut thickness of the octet lattice is 1.27 μm (Fig. 1G). The minimal wall thickness of SP lattices is 850 nm (Fig. 1I). Both Au_{25} and $Ag_{28}Pt$ nanoclusters can also function as two-photon acid generators for cationic polymerization of epoxy monomers (EEC) with scanning speeds up to 100 mm/s (fig. S6).

The $Ag_{28}Pt$ and Au_{25} nanoclusters show state-of-the-art performance comparable to the most efficient organic two-photon initiators. The efficiency of different photoresists can be compared using a dimensionless figure of merit (FOM) (see fig. S7 for details). Kiefer *et al.* (*13*) reported that only six types of organic photoresists have a FOM ≥ 10^2 at velocities ≥ 10 mm/s (fig. S7). Our nanocluster-PETA photoresists have a FOM ≥ 10^2 at velocities up to 100 mm/s. The minimum required concentration of the nanocluster in resin is 0.47 to 2.75 μmol/g (details in SI), which is among the lowest compared with hundreds of organic two-photon initiators (*13*, *14*). In addition, the nanocluster photoresists can be applied to more than one class of monomer, which provides an elegant strategy for printing high-quality nanocomposites.

The mechanical performance of the printed nanocluster-polymer structures is evaluated with in situ SEM compression testing (Fig. 2). The 8 wt % Au_{25} PETA photoresist was used to fabricate cylindrical pillars 2.5 μm in radius and 10 μm in height. Engineering stress-strain curves show a nonlinear loading curve with significant hardening at strains above ~20% (fig. S8). This strain hardening is still apparent when corrected for true stress and strain (fig. S8). Sudden failure occurs at 1.0 ± 0.2 GPa and a strain of ~50%, although crack initiation was observed at ~40% strain in some pillars (fig. S8 and movie S3). The elastic modulus was found to be 3.7 ± 0.2 GPa (fig. S9). Cyclic tests showed 70% recovery for samples that were loaded to 30% strain (Fig. 2A, fig. S10, and movie S4). The high strength, stiffness, and strain hardening lead to an energy absorption of 110 ± 50 MJ/m^3 before initial crack formation. This mechanical behavior is also apparent in honeycomb structures fabricated using 10 wt % $Ag_{28}Pt$ PETA photoresist (Fig. 2B). The honeycombs have a height of 8.2 ± 0.1 μm, cell side length of 2.6 ± 0.2 μm, and wall thickness of 800 ± 50 nm (Fig. 2C). This results in a density

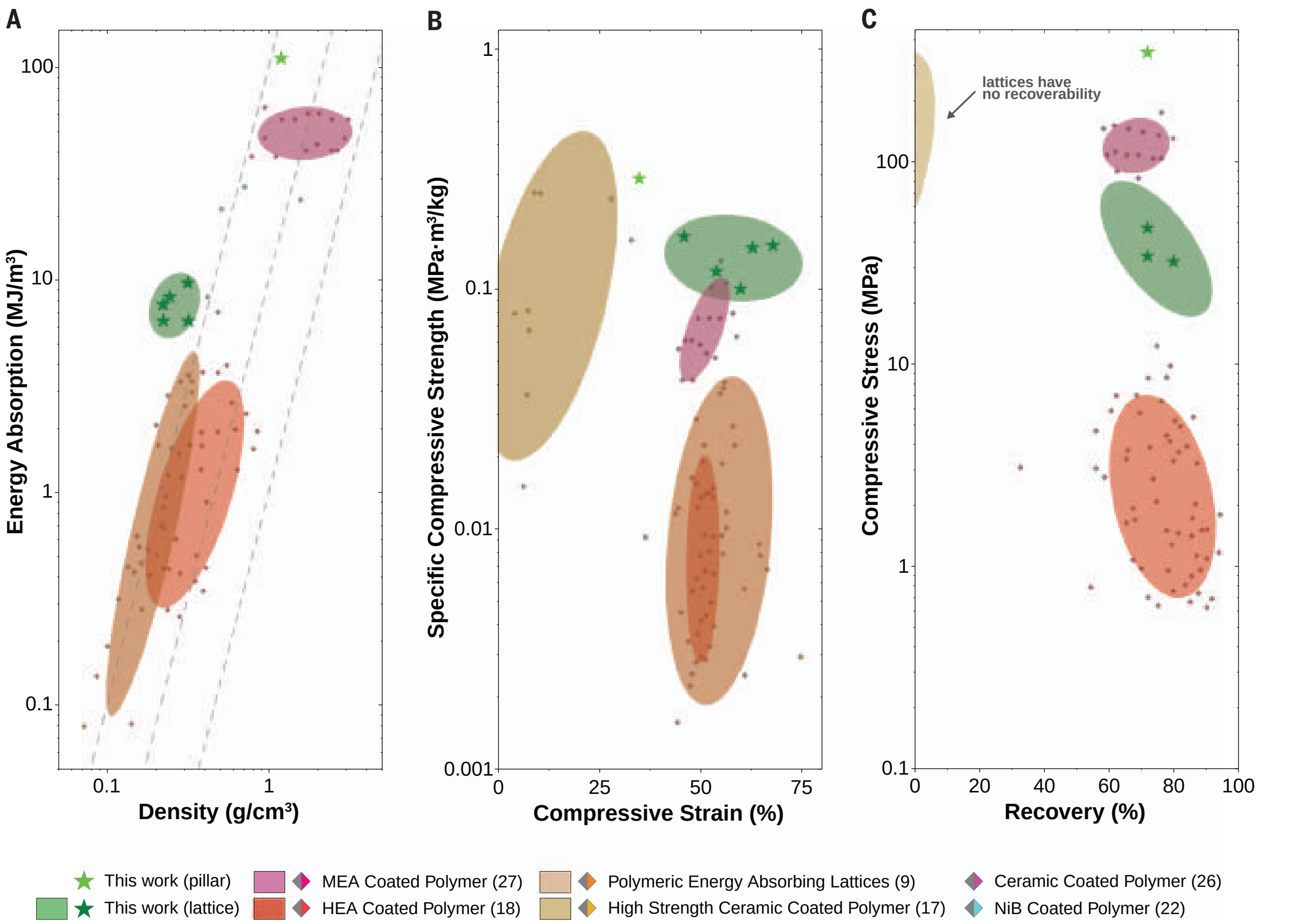


Fig. 3. Comparison of mechanical performance. (**A**) Energy absorption versus density, (**B**) specific compressive strength versus compressive strain, and (**C**) compressive stress versus recovery of the nanocluster-polymer nanolattices and pillars compared with other micro- and nanolattices (i.e., inorganic coatings on polymer lattices). Data are from (*9*, *17*, *18*, *22*, *26*, *27*).

of 0.56 g/cm^3 and a relative density of ~48%. The compressive stress-strain response shows a linear region followed by a distinct yielding event at 22.1 ± 2.1 MPa (Fig. 2B). After this, the stress-strain curve is nonlinear, with significant hardening at large strains until ultimate failure at ~25% strain. The honeycombs have a nominal failure stress of 178 ± 52 MPa. In comparison, polymeric honeycombs (e.g., commercial Ip-Dip) with similar structural density generally do not show such significant hardening beyond the yield point (*22*) (Fig. 2B). Thus, our nanocluster-polymer structure results in far better ultimate strength and energy absorption.

We examine the origin of the nanocluster composite mechanical behavior. By using the Halpin-Tsai model, we determine that the nanocluster composites are not merely mechanical reinforcements but perturb the bonding and physical properties of the polymer matrix in an interphase region (see supplementary materials for detailed calculations). This interphase region is estimated to have a thickness of ~1 nm (0.25 of the volume fraction of the nanocluster composite) and an elastic modulus of 24 GPa. This is reasonable given the ~5 nm average spacing of the nanoclusters. The interphase volume is roughly three times larger than that of the nanoclusters. We hypothesize that the nanocluster composite deformation mechanism is similar to that of thermoplastic copolymers (e.g., thermoplastic polyurethane) (*23*) or hybrid materials with interpenetrating networks (*24*). These thermoplastics have similar stress-strain behavior, including strain hardening and hysteresis between loading and unloading. The average spacing of the nanoclusters is on par with the spacing of hard domains in thermoplastic copolymers. The interactions of the nanoclusters may be similar to the alignment of the hard domains in thermoplastic copolymers, which leads to high strength, stiffness, and significant strain hardening despite the lack of chemical bonding.

Octet and Schwarz primitive (SP) lattices are fabricated from 8 wt % Au_{25} photoresist. These structures show high energy absorption before densification. Octet lattices with relative densities of 19 and 27% and SP lattices with relative densities of 20 and 26% have energy absorption capacities up to 7.6 MJ/m^3 and 9.7 MJ/m^3, respectively. In situ SEM compression testing reveals that this best-of-class performance is due to the characteristic material properties of the nanocluster-polymer composite (Fig. 2, D to K, and movies S5 and S6). Low-density cellular lattices typically collapse in a layer-by-layer manner, which leads to regions of negative stiffness in the plateau portion of the stress-strain curve (fig. S11) (*18*, *22*, *25*). This results in reduced energy absorption capacity. Layer-by-layer collapse and negative stiffness are suppressed in the nanocluster-polymer nanolattices, which retain high recoverability of up to 80% even after compression to ~60% strain. Tensegrity structures have been constructed to promote uniform deformation in cellular polymeric lattices (*25*). We show that such complex lattice architectures are not always necessary to achieve optimized behavior.

The specific energy absorption of the nanocluster-polymer nanolattices and pillars is better than polymer micro- and nanolattices with inorganic coatings (*9*, *17*, *18*, *22*, *26*, *27*), as well as conventional material systems (Fig. 3A).

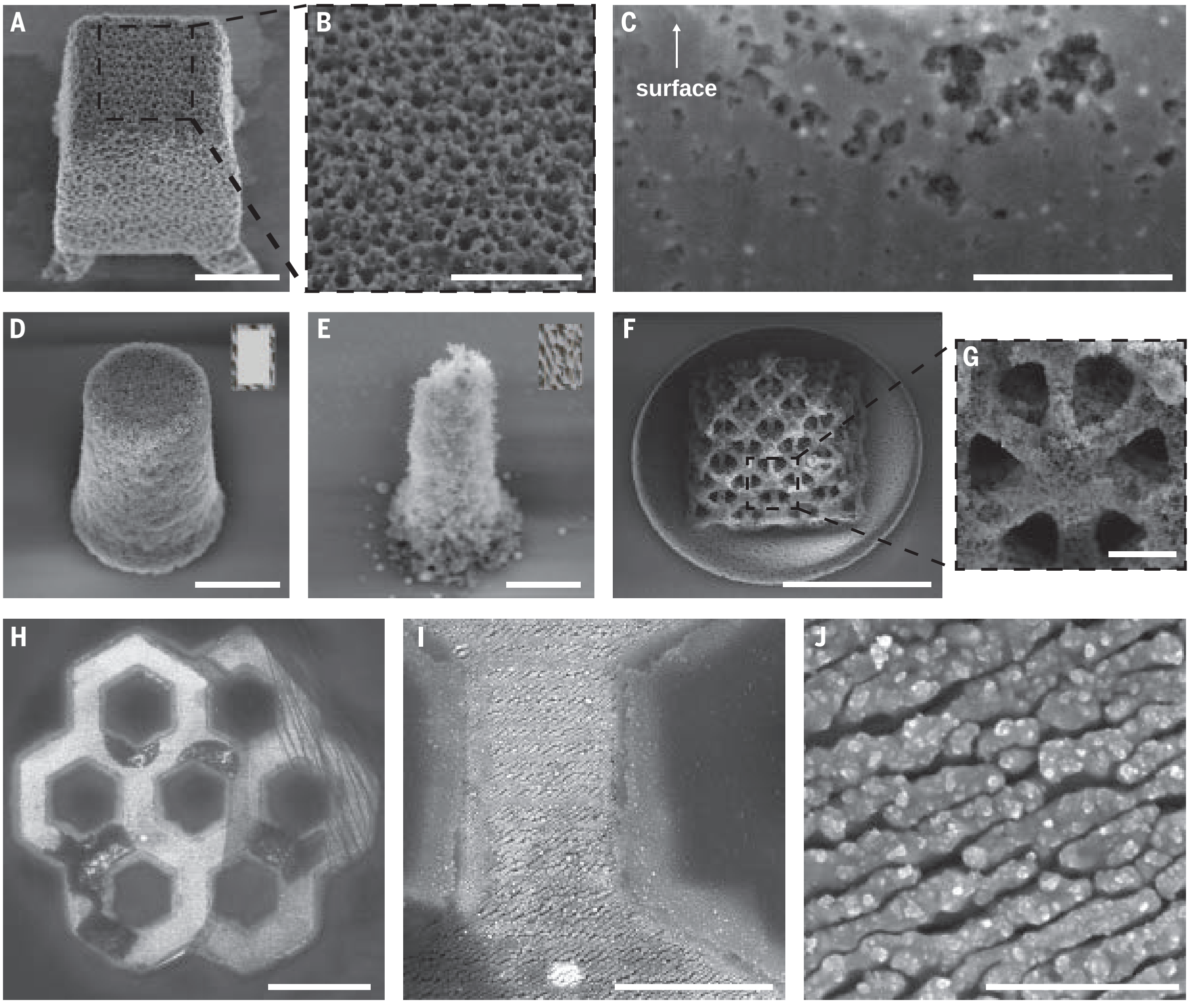


Fig. 4. Hierarchical, tunable, and anisotropic nanoporous structures of glassy carbon and silk protein. (**A**) Nanoporous cube on supports. (**B**) Magnified view of top surface. (**C**) Cross-sectional image of nanoporous cube. (**D**) Nanoporous shell-solid core pillar. (**E**) Nanoporous pillar. (**F** and **G**) Nanoporous octet lattice. (**H** to **J**) Silk fibroin honeycomb structures at different length scales. Scale bars: (A) 2 μm, (B) 1 μm, (C) 200 nm, (D) 5 μm (E) 1 μm, (F) 20 μm (G) 2 μm, (H) 10 μm, (I) 3 μm, (J) 500 nm.

The nanocluster composites also have high specific compressive strength at large strains (Fig. 3B) and high recoverability at large compressive stresses (Fig. 3C). Glassy carbon structures, although considerably strong and stiff (*6*), are not included in this comparison because of their brittle and nonrecoverable behavior. By using nanocluster-based photoresists, these mechanical properties can be achieved in a single printing step, in contrast to the multiple fabrication steps required for core-shell composite lattices.

Nanocluster-based photoresists can be used to fabricate complex nanoporous structures (Fig. 4). Nanocluster-polymer composites were transformed into glassy carbon with complex nanoporous features through pyrolysis at 900°C under argon flow. The presence of glassy carbon was confirmed by Raman analysis (fig. S12). Figure 4A shows a nanoporous cube printed using 20 wt % $Ag_{28}Pt$ photoresist. The average pore size and porosity on the surface of the cube are ~56 ± 23 nm (Fig. 4B and fig. S13) and ~50%, respectively. The nanopores extend ~500 to 700 nm into the structure (Fig. 4C) and decrease in size and density further from the surface. Thus, larger structures such as the ~8 μm diameter pillar shown in Fig. 4D have graded porosity that consists of a solid core surrounded by a nanoporous shell. Small structures <2 μm in size are porous throughout the structure (Fig. 4E). Pyrolyzed octet lattices have two levels of porosity: geometrically designed spaces between lattice struts and nanocluster-induced nanopores within lattice struts (Fig. 4, F and G).

Printed structures containing fewer nanoclusters result in decreased porosity after pyrolysis (fig. S14). Previous work involving pyrolysis of PETA-based photoresists resulted in glassy carbon without porosity (*6*). We propose that the nanopores result from the melting of nanoclusters during pyrolysis. As the structure is heated to 900°C, the nanoclusters merge to form larger particles and eventually precipitate out of the structure. The melting temperature of 1- to 2-nm nanoclusters is estimated to be 600 to 900°C (*28*), which is consistent with the observation that nanoporous glassy carbon forms only above 550°C (fig. S14). This is also supported by the observation of ~100- to 500-nm metal particles on the surface of some pyrolyzed structures and the nearby substrate (fig. S15). Energy-dispersive x-ray spectroscopy (EDS) confirmed the presence of Au or Ag in or on the pyrolyzed structures (fig. S16). Nanoporous glassy carbon pillars with a density of ~1.07 g/cm^3 were found to have a compressive failure strength of 1.3 ± 0.9 GPa (figs. S17 and S18).

It is challenging to fabricate hierarchical structures with anisotropic pores, although this is a common motif in natural structural materials (e.g., aligned collagen and pores in bone) (*10*). In addition, the two-photon lithography of protein structures is generally

slow, with reported speeds in the range of 1 to 1000 μm/s (*29*). Here, a nanocluster-protein photoresist (detailed composition in the SM) is developed to address these issues. This photoresist exploits the efficient singlet oxygen generation and significant photothermal effect of metal nanoclusters under excitation (*30*). This induces photo-crosslinking of proteins through oxidization of the tyrosine residues (Fig. 1B) (*31*), as well as the formation of directional β-sheet crystalline regions though local heating. A water-dispersible Au_{22} nanocluster (*32*) was synthesized and used in the photoresist. Silk fibroin structures are printed at speeds up to 100 mm/s (laser power ~25 mW). The printed protein structures are composed of aligned bundles (Fig. 4, H to J), which indicates that directional self-assembly occurred during fabrication. Raman spectroscopy confirms the formation of aligned crystalline domains made of β-sheets (fig. S19). The anisotropic nanovoids (fig. S20) between these domains may be due to the destruction of softer parts of the protein chains by the strong photothermal effect.

In summary, this work develops metal nanocluster-based photoresists that are used to fabricate nanocluster-polymer nanolattices, as well as nanoporous glassy carbon and protein structures with unprecedented structural complexity. We show that nanoclusters are versatile and highly efficient two-photon activators that are applicable across different classes of monomers. The nanocluster-polymer nanolattices show strain hardening behavior, which leads to a combination of high specific energy absorption, specific strength, deformability, and recoverability. Based on this simple and general framework, there is a broad opportunity to directly print additional mechanical metamaterials by combining the hundreds of available metal nanoclusters with different types of monomers and rationally designed 3D topologies.

REFERENCES AND NOTES

1. S. K. Saha *et al*., *Science* **366**, 105–109 (2019).
2. S. Kawata, H.-B. Sun, T. Tanaka, K. Takada, *Nature* **412**, 697–698 (2001).
3. V. Hahn *et al*., *Nat. Photonics* **15**, 932–938 (2021).
4. Z. C. Eckel *et al*., *Science* **351**, 58–62 (2016).
5. J. Bauer *et al*., *Adv. Mater.* **29**, 1701850 (2017).
6. X. Zhang, A. Vyatskikh, H. Gao, J. R. Greer, X. Li, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 6665–6672 (2019).
7. L. R. Meza, S. Das, J. R. Greer, *Science* **345**, 1322–1326 (2014).
8. X. Zheng *et al*., *Science* **344**, 1373–1377 (2014).
9. S. Yuan, C. K. Chua, K. Zhou, *Adv. Mater. Technol.* **4**, 1800419 (2019).
10. O. A. Tertuliano, J. R. Greer, *Nat. Mater.* **15**, 1195–1202 (2016).
11. W. Zhou *et al*., *Science* **296**, 1106–1109 (2002).
12. L. Li, R. R. Gattass, E. Gershgoren, H. Hwang, J. T. Fourkas, *Science* **324**, 910–913 (2009).
13. P. Kiefer *et al*., *Adv. Opt. Mater.* **8**, 2000895 (2020).
14. T. Wloka, M. Gottschaldt, U. S. Schubert, *Chem. Eur. J.* **28**, e202104191 (2022).
15. M. Carlotti, V. Mattoli, *Small* **15**, 1902687 (2019).
16. C. M. Spadaccini, in *Three-Dimensional Microfabrication Using Two-Photon Polymerization (Second Edition)*, T. Baldacchini, Ed. (William Andrew Publishing, 2020), Chapter 13.4
17. J. Bauer, S. Hengsbach, I. Tesari, R. Schwaiger, O. Kraft, . *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2453–2458 (2014).
18. X. Zhang *et al*., *Nano Lett.* **18**, 4247–4256 (2018).
19. R. Jin, C. Zeng, M. Zhou, Y. Chen, *Chem. Rev.* **116**, 10346–10413 (2016).
20. X. Lin *et al*., *Nano Res.* **12**, 309–314 (2019).
21. Q. Li, C. J. Zeman4th, Z. Ma, G. C. Schatz, X. W. Gu, *Small* **17**, 2007992 (2021).
22. M. Mieszala *et al*., *Small* **13**, 1602514 (2017).
23. H. J. Qi, M. C. Boyce, *Mech. Mater.* **37**, 817–839 (2005).
24. B. M. Novak, *Adv. Mater.* **5**, 422–433 (1993).
25. J. Bauer, J. A. Kraus, C. Crook, J. J. Rimoli, L. Valdevit, *Adv. Mater.* **33**, 2005647 (2021).
26. A. Schroer, J. M. Wheeler, R. Schwaiger, *J. Mater. Res.* **33**, 274–289 (2018).
27. X. Feng *et al*., *Mater. Today* **42**, 10–16 (2021).
28. F. Gao, Z. Gu, in *Handbook of Nanoparticles*, M. Aliofkhazraei, Ed. (Springer, 2016), pp. 661–690.
29. D. Serien, K. Sugioka, *OEA* **1**, 180008–180018 (2018).
30. Y. Xie, W. Zheng, X. Jiang, *ACS Appl. Mater. Interfaces* **12**, 9041–9049 (2020).
31. X. Mu, J. K. Sahoo, P. Cebe, D. L. Kaplan, *Polymers* **12**, 2936 (2020).
32. K. Pyo *et al*., *J. Am. Chem. Soc.* **137**, 8244–8250 (2015).
33. Q. Li *et al*., Mechanical nanolattices printed using nanocluster-based photoresists, Dryad (2022); https://doi.org/10.5061/dryad.ksn02v77x.

ACKNOWLEDGMENTS

This work was financially supported by the National Science Foundation (CMMI-2052251 and DMR-2002936/2002891) and the American Chemical Society Petroleum Research Doctoral New Investigator Grant. D.D. acknowledges the National Science Foundation Graduate Research Fellowship under grant 1656518. J.K. was supported by a Stanford Graduate Fellowship. Part of this work was performed at the Stanford Nano Shared Facilities (SNSF) and Stanford Nanofabrication Facility (SNF), which are supported by the National Science Foundation under award ECCS-1542152. **Author contributions:** Q.L. conceived the initial idea. X.W.G. and Q.L. designed the research. J.K., D.D., and Q.L. performed 3D printing. J.K., D.D., and M.M.W. conducted mechanical tests. D.D., J.K., and O.A.T. performed structural characterization. C.J.Z. and G.C.S. performed simulations. Q.L., J.K., D.D., O.A.T., C.J.Z., and X.W.G. wrote the paper, and all authors commented on it. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data are available in the manuscript or the supplementary material or are deposited in Dryad (*33*).

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo6997
Materials and Methods
Figs. S1 to S20
References (*34–39*)
Movies S1 to S6

Submitted 19 February 2022; resubmitted 12 September 2022
Accepted 20 October 2022
10.1126/science.abo6997

ORGANIC CHEMISTRY

Halogenation of the 3-position of pyridines through Zincke imine intermediates

Benjamin T. Boyle†, Jeffrey N. Levy†, Louis de Lescure, Robert S. Paton*, Andrew McNally*

Pyridine halogenation reactions are crucial for obtaining the vast array of derivatives required for drug and agrochemical development. However, despite more than a century of synthetic endeavors, halogenation processes that selectively functionalize the carbon–hydrogen bond in the 3-position of a broad range of pyridine precursors remain largely elusive. We report a reaction sequence of pyridyl ring opening, halogenation, and ring closing whereby the acyclic Zincke imine intermediates undergo highly regioselective halogenation reactions under mild conditions. Experimental and computational mechanistic studies indicate that the nature of the halogen electrophile can modify the selectivity-determining step. Using this method, we produced a diverse set of 3-halopyridines and demonstrated late-stage halogenation of complex pharmaceuticals and agrochemicals.

Pyridines are widely found in biologically relevant molecules, and their prevalence results from an interplay of the heterocycle's intrinsic properties and substituents (*1–4*). Therefore, regioselective pyridine C–H halogenation reactions are vital because installing a carbon–halogen (C–Hal) bond enables numerous subsequent bond-forming reactions (Fig. 1A). In pharmaceutical and agrochemical research, halopyridines are key intermediates to diversify candidate compounds for structure-activity relationship studies as well as target-orientated synthesis (*5–10*). Halopyridines are also inherently valuable in bioactive compounds such as etoricoxib. However, despite reports of pyridine halogenation reactions dating back to the late 19th century, there are still major limitations to this valuable bond construction (*11–13*). In particular, processes selective for the 3-position tend to be incompatible with the functionality of many substrates of interest for their bioactivity (*14*).

Pyridine halogenation reactions using electrophilic aromatic substitution are electronically mismatched processes that require harsh conditions (*15*, *16*). Specifically, elemental halides with strong Brønsted or Lewis acids are often used at elevated temperatures to compensate for the poor π nucleophilicity of pyridine rings (*5*, *6*). Substrate scope is limited, and despite being 3-selective, reactions often result in regioisomeric mixtures. Metalation-halogenation

Department of Chemistry, Colorado State University, Fort Collins, CO 80523, USA.
***Corresponding author. Email: andy.mcnally@colostate.edu (A.M.); robert.paton@colostate.edu (R.S.P.)**
†These authors contributed equally to this work.

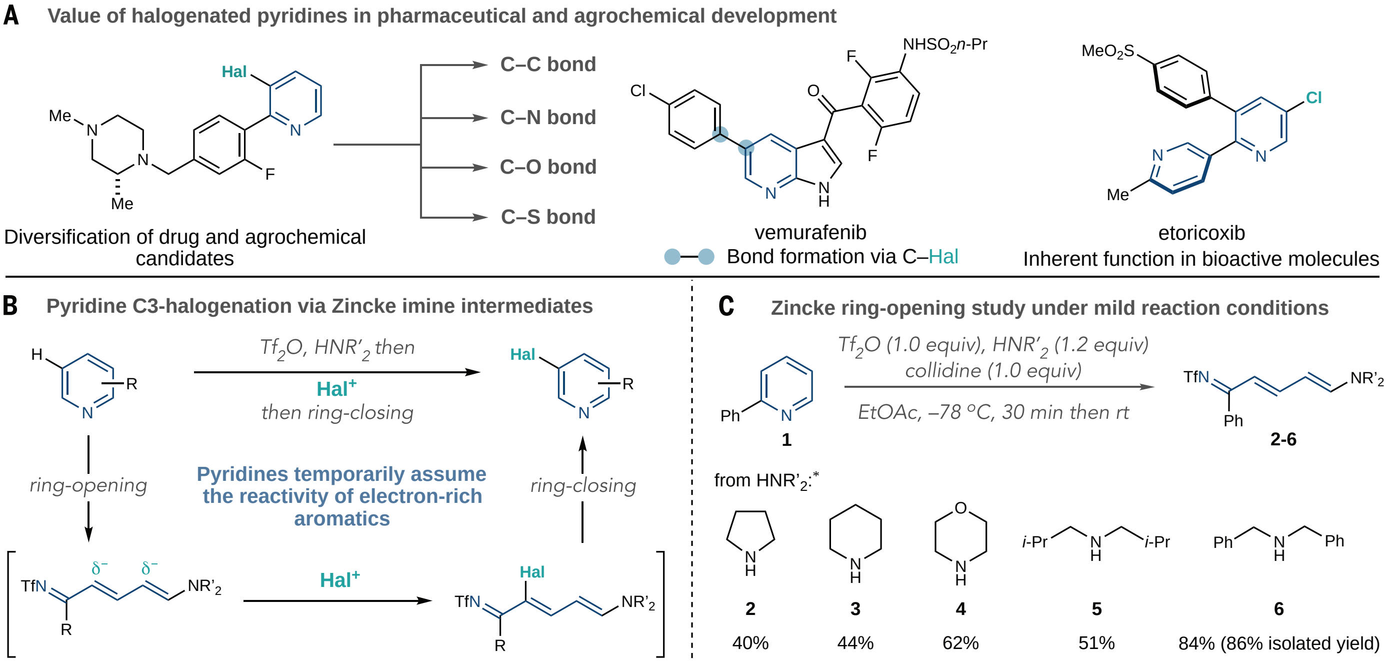


Fig. 1. Importance of pyridine halogenation reactions and a distinct strategy based on ring-opened intermediates. (**A**) Examples of pyridine halogenation products and derivatives in drug and agrochemical development. (**B**) A ring-opening, halogenation, ring-closing synthetic strategy. (**C**) Ring-opening study using amine nucleophiles. *Yields calculated from integrated ^{1}H NMR spectra of the crude reaction mixture using triphenylmethane as an internal standard. R, general organic group; Me, methyl; Hal, halogen; Tf, trifluoromethylsulfonyl; collidine, 2,4,6-trimethylpyridine; Et, ethyl; Ac, acetate; Ph, phenyl; *i*-Pr, isopropyl; rt, room temperature.

sequences with strong bases are another approach for pyridine halogenation, but most require directing groups to access the 3-position reliably (*17–19*). Beyond these two approaches, there has been little progress over the past century, so the synthetic community turned to other versatile functional groups in place of halides. Iridium-catalyzed borylations and silylations are the most notable processes and are 3-selective for certain classes of substituted pyridines or through control with ligand architectures (*20–26*). Here, we present an alternative approach for pyridine halogenation using a ring-opening, halogenation, ring-closing strategy (Fig. 1B). This "one-pot" protocol uses a modified version of the classic Zincke ring-opening reaction that converts pyridines into azatriene intermediates, or "Zincke imines." This synthetic maneuver temporarily transforms pyridines from electron-deficient heterocycles to a series of polarized alkenes that that undergo electrophilic substitution reactions, much like electron-rich aromatics. The halogenation process is 3-selective across a range of substituted pyridines and is functionally compatible with complex structures of pharmaceutical interest.

Current methods for Zincke ring-opening chemistry cannot enable the general process for pyridine halogenation described in Fig. 1B (*27–29*). The pyridine *N*-activation step requires forcing conditions and often fails when 2-position substituents are present. This limitation excludes many classes of substituted pyridines. To address these restrictions, we focused on ring opening of *N*Tf-pyridinium salts. These reactive intermediates are readily formed from the pyridine and Tf_2O at low temperatures and encompass most substitution patterns except 2,6-disubstitution. Toscano *et al.* reported a ring opening with Tf_2O, but did not extend this process beyond pyridine, and observed mixtures of ring-opened products (*30*). With 2-phenylpyridine (**1**), we surveyed a series of aliphatic amines as nucleophiles (Fig. 1C). Pyrrolidine, piperidine, and morpholine gave moderate yields of ring-opened products, as did diisobutylamine (**2** to **5**). However, dibenzylamine proved optimal, with **6** formed in high yield. Tables S1 and S2 show further studies of pyridine ring openings using this protocol. We anticipated that the ring closure of Zincke imines such as **6** would be straightforward based on precedent in related Zincke systems (*31*). Examples typically use ammonium salts or Brønsted acids to obtain pyridines, and these reagents could potentially facilitate the one-pot sequence outlined in Fig. 1B (see below).

For the next stage of reaction development, we tested Zincke imine **6** with halogen electrophiles (Fig. 2A). At the outset, we did not expect that C–Hal bond formation would be selective because there was no clear electronic bias between the two δ^- sites at C3 and C5 (see below) in **6** (*32*). Toscano *et al.* did observe a 3-selective reaction between an unsubstituted *N*Tf-Zincke imine derived from pyridine and a pyridinium electrophile; however, there were no investigations into Zincke imines derived from substituted pyridines (*30*). When we subjected **6** to *N*-iodosuccinimide (NIS) at room temperature for 5 min, we observed **7-I** in 92% yield and >20:1 3- versus 5-selectivity (**8-I**) based on analysis of the crude ^{1}H nuclear magnetic resonance (NMR) spectrum. Using *N*-bromosuccinimide (NBS), we observed a 4.4:1 ratio of **7-Br** and **8-Br** at room temperature and >20:1 selectivity and 92% of **7-Br** at −78°C in CH_2Cl_2. Chlorinating with *N*-chlorosuccinimide (NCS) was lower yielding and less selective (**7-Cl** and **8-Cl**), and it decomposed **6**. At this point, we tested conditions for ring closure and found that heating **8-I** in a EtOAc/EtOH solvent mixture at 60°C with 10 equivalents of ammonium acetate smoothly formed 3-iodopyridine **9-I**. Table S10 shows further studies of this ring-closing step.

Halogenating 3-substituted imines (Fig. 2B) required a different procedure. Ring opening of 3-phenylpyridine produced Zincke imine **10** in 82% yield, but NIS did not react with **10** at room temperature and decomposed to unidentified by-products when heated to 50°C. However, when we included two equivalents of trifluoroacetic acid (TFA), iodopyridine **12-I** formed in high yield (*33*). We attribute this in situ rearomatization to steric interactions between the 3- and 5-substituents in **11-I**; TFA can promote the alkene isomerization-recyclization sequence,

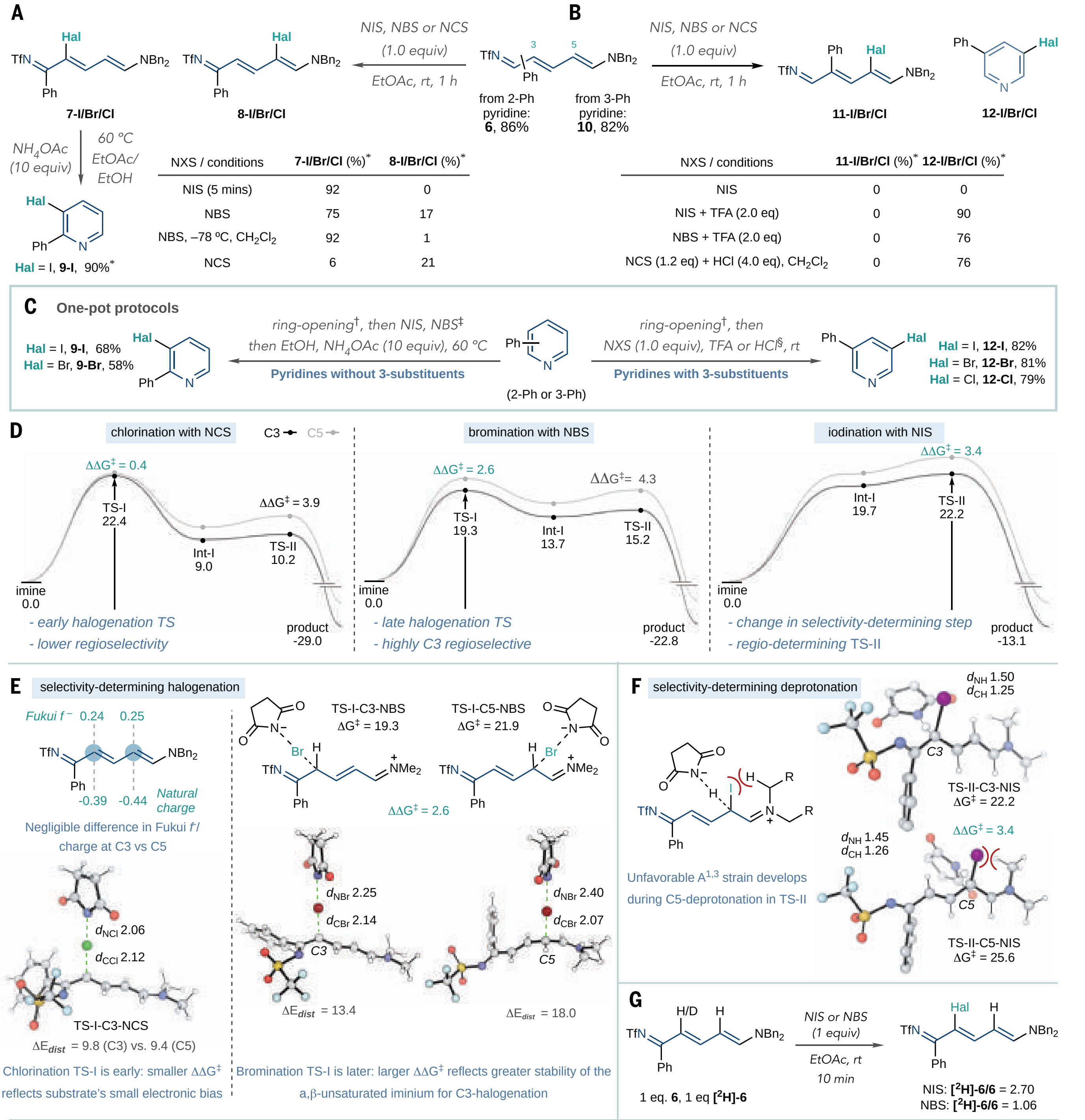


Fig. 2. Reaction development and mechanistic investigation. (**A**) Halogenation of 2-substituted Zincke imines and ring closure using ammonium salts. *Yields calculated from [1]H NMR of the crude reaction mixture using triphenylmethane as an internal standard. (**B**) Halogenation of 3-substituted Zincke imines. (**C**) One-pot pyridine halogenation protocols. †Ring-opening conditions: Tf_2O (1.0 equiv), $HNBn_2$ (1.2 equiv), collidine (1.0 equiv), EtOAc, –78°C, 30 min then warm to rt. ‡Ring-opening and halogenation steps performed in CH_2Cl_2 instead of EtOAc when using NBS. §Halogenation–ring-closure protocol for pyridines with 3-substituents: NIS (1.0 equiv) or NBS (1.0 equiv) and TFA (2.0 equiv). NCS (1.0 equiv) and HCl (4.0 equiv) and ring opening were performed in CH_2Cl_2. Isolated yields are shown. (**D**) Computed Gibbs energy profiles (in kilocalories per mole) for imine halogenation by *N*-halosuccinimides. (**E**) Analysis of selectivity-determining chlorination and bromination transition structures **TS-I**. (**F**) Analysis of selectivity-determining deprotonation transition structure **TS-II** for iodination. (**G**). Isotope competition experiment. Bn, benzyl group; NXS, *N*-halosuccinimides.

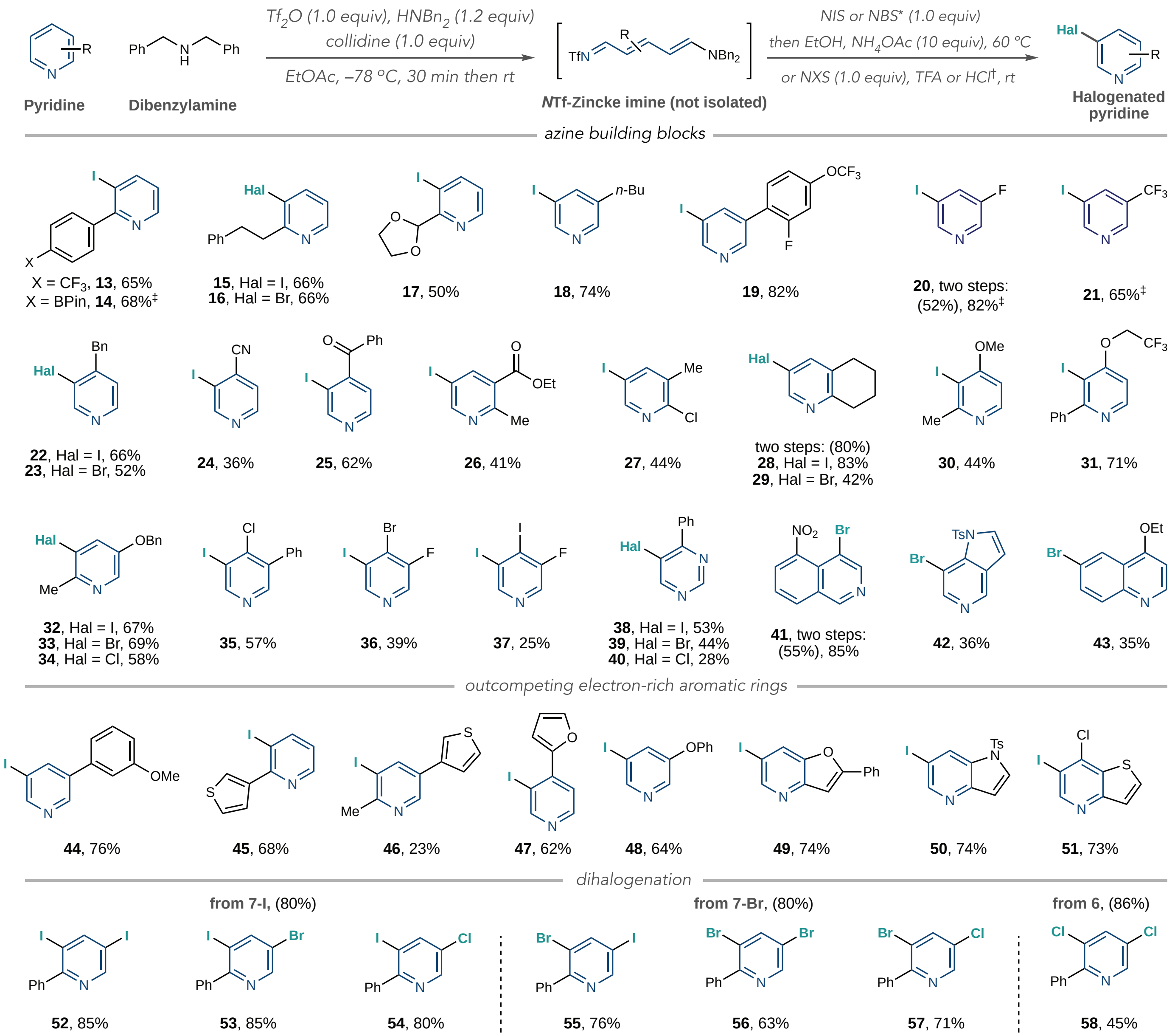


Fig. 3. One-pot pyridine halogenation: scope of building-block pyridines and dihalogenation reactions. Yields of isolated products are shown. Yields in parentheses are those of isolated Zincke imines. *Ring-opening and halogenation steps performed in CH_2Cl_2 instead of EtOAc when using NBS. †Halogenation–ring-closure protocol for pyridines with 3-substituents: NIS (1.0 equiv) or NBS (1.0 equiv) and TFA (2.0 equiv). NCS (1.0 equiv) and HCl (4.0 equiv), and ring opening is performed in CH_2Cl_2. ‡^{1}H NMR yields using triphenylmethane as an internal standard. Pin, pinacolato; *n*-Bu, normal butyl group; Ts, tosyl.

as well as the subsequent elimination of dibenzylamine and N–S bond cleavage of the resulting *N*Tf-pyridinium salt to reform the pyridine ring system. Brominating with NBS and TFA (**12-Br**) resulted in the same outcome, and a modified procedure using NCS with HCl in CH_2Cl_2 chlorinated **10** effectively to form **12-Cl**. The results (Fig. 2, A and B) provide two halogenation protocols to halogenate *N*Tf-Zincke imine intermediates depending on their substitution pattern; further details of this optimization study are described in tables S3 to S8.

Next, we combined the ring-opening, halogenation, and ring-closing steps into a one-pot process (Fig. 2C). For pyridines without 3-substituents, such as 2-phenylpyridine, ring opening using dibenzylamine and halogenation with NIS or NBS (Fig. 2A) was appropriate. Adding NH_4OAc and EtOH to the reaction mixture and heating to 60°C induced ring closure to reform pyridines **9-I** and **9-Br**. Adding an equivalent of trimethoxybenzene after the halogenation step proved helpful in quenching any remaining *N*-halosuccinimide before cyclization. When 3-substituents were present, such as in 3-phenylpyridine, ring opening followed by addition of the *N*-halosuccinimides and acids described in Fig. 2B comprised the one-pot process (**12-I, 12-Br**, and **12-Cl**). These protocols do not require intermediate workup or purification steps and involve adding reagents sequentially to the same reaction vessel.

We then studied the mechanism and regioselectivity of Zincke imine halogenation by *N*-halosuccinimides with quantum chemistry at the B3LYP-D3(BJ)/def2-TZVP//ωB97X-D/6-31+G(d,p) level of theory with a "solvation model based on density (SMD)" description of CH_2Cl_2. The most favorable pathway for halogenating **6** (Fig. 2D), with the NBn_2 portion simplified to NMe_2, consisted of electrophilic

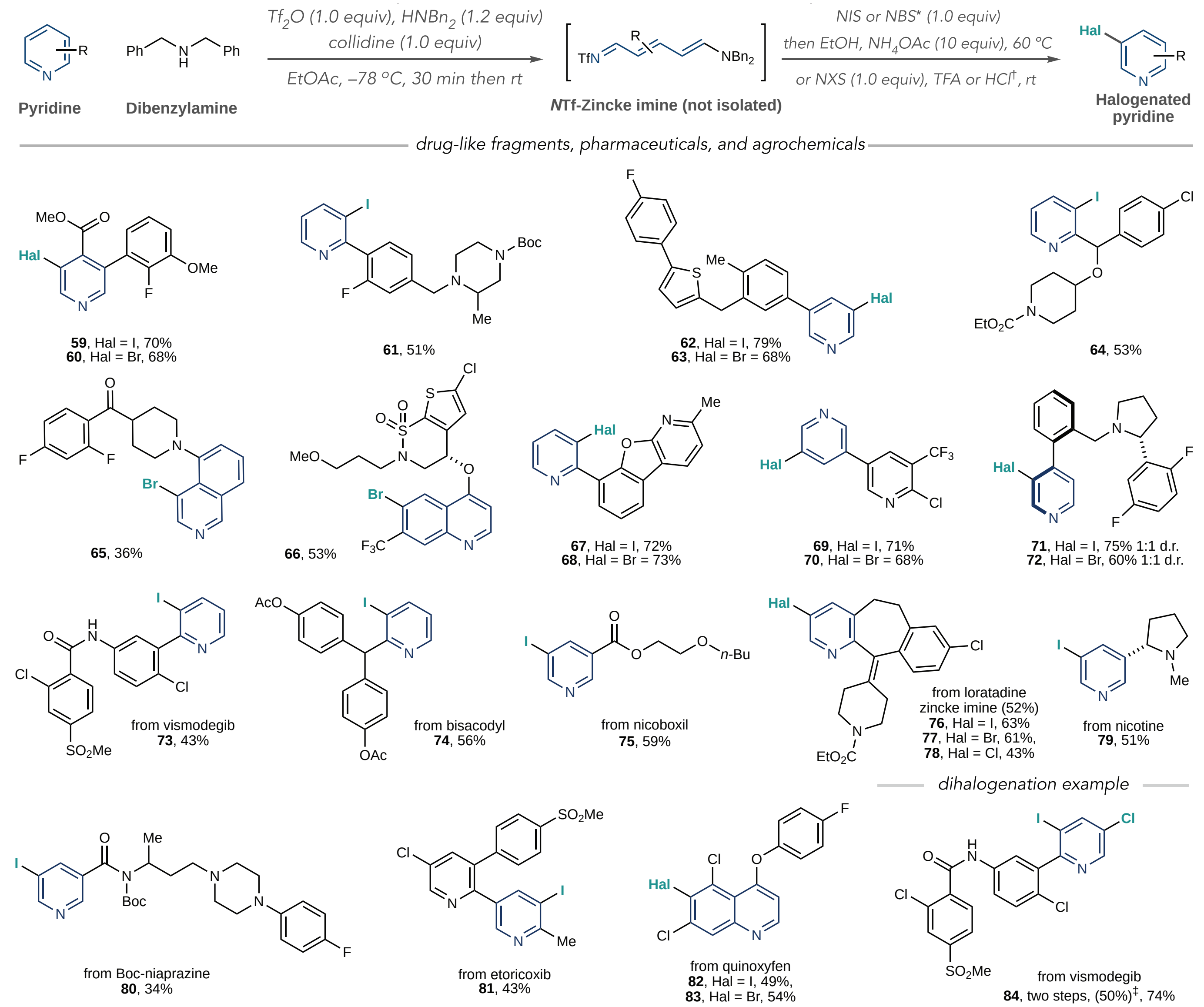


Fig. 4. One-pot halogenation of complex azines. Yields of isolated products are shown. Yields in parentheses are those of isolated Zincke imines. *Ring-opening and halogenation steps performed in CH_2Cl_2 instead of EtOAc when using NBS. †Halogenation–ring-closure protocol for pyridines with 3-substituents: NIS (1.0 equiv) or NBS (1.0 equiv) and TFA (2.0 equiv). NCS (1.0 equiv) and HCl (4.0 equiv), and ring opening is performed in CH_2Cl_2. ‡Isolated yield of iodinated Zincke imine. Boc, *tert*-butoxycarbonyl group.

addition (**TS-I**) followed by deprotonation (**TS-II**). Calculations showed that outer-sphere electron transfer processes were inaccessible ($\Delta G > 34$ kcal mol^{-1}); figs. S11 to S23 provide full computational details.

Regioselectivity in electrophilic halogenation is conventionally understood in terms of differences in frontier orbital coefficients, atomic charges, or nucleophilicity parameters (*34*). However, the electronic environments at the C3 and C5 positions of Zincke imine **6** are very similar in terms of Fukui *f* coefficients (0.24 versus 0.25), natural charges (–0.39 versus –0.44), and HOMO coefficients (both 0.26), necessitating an alternative rationale for the high levels of C3 selectivity in this transformation (Fig. 2E). Computed energy profiles (Fig. 2D) for **6** reacting with NCS, NBS, and NIS suggest an irreversible overall reaction governed by kinetically controlled regioselectivity in each case, with $\Delta\Delta G^{\ddagger}$ values in quantitative agreement with experiment. Breaking the extended conjugation in the Zincke imine to form positively charged s-complex **Int-I** is endergonic: G_{rel} values of this high-energy intermediate increase for C–Cl, C–Br, and C–I formation, respectively (9.0, 13.7, and 19.7 kcal mol^{-1}, respectively). These differences give rise to two distinct selectivity-determining regimes: an irreversible C–Hal bond-forming step (**TS-I**) determines regioselectivity for chlorination and bromination, whereas C–I bond formation is reversible, and the second deprotonation step (**TS-II**) dictates the regioselectivity (*35*).

The higher C3 selectivity for bromination over chlorination ($\Delta\Delta G^{\ddagger}$ of 2.5 and 0.4 kcal mol^{-1}, respectively) can be understood on the basis of early versus late transition state arguments. Chlorination results in a more stable intermediate than bromination and, in accordance with the Hammond postulate, proceeds through an earlier TS [as shown by smaller TS distortion energies in Fig. 2E and smaller pyramidalization values at the reacting carbon atom (fig. S19)]. Given the electronic similarity between the C3 and C5 positions in the Zincke imine and the absence of appreciable intermolecular noncovalent interactions (fig. S17), there is little regioisomeric preference in the early chlorination TS. Bromination proceeds through a later TS that incurs larger substrate

distortion and strongly favors addition at C3 ($\Delta\Delta E_{dist}$ = 4.6 kcal mol^{-1}). This difference in distortion energies reflects a more stable developing a,b–unsaturated iminium ion in **TS-I-C3-NBS** versus **TS-I-C5-NBS** ($\Delta\Delta G^{\ddagger}$ = 2.6 kcal mol^{-1}). By contrast, the high regioselectivity of iodination ($\Delta\Delta G^{\ddagger}$ = 3.4 kcal mol^{-1}) results from a rate- and selectivity-determining deprotonation (**TS-II**). Restoring planarity in **TS-II-C5-NIS** results in increased $A^{1,3}$ strain between the enamine carbon substituent and the iodide (Fig. 2F). The bond and dihedral angles in the C5-iodinated product clearly reflect this developing strain (fig. S21).

Intermolecular competition experiments validate these computational predictions (Fig. 2G). We reacted one equivalent each of **6** and **[^{2}H]-6** until this mixture consumed one equivalent of NIS. The **[^{2}H]-6/6** ratio of 2.70 measured at the end of the reaction indicates that **6** is consumed preferentially and supports the selectivity-determining deprotonation predicted by theory (Fig. 2F). The same experiment using NBS did not show a significant bias between **[^{2}H]-6** and **6** at the end of the reaction, consistent with selectivity-determining C–Br bond formation.

Next, we applied a series of azine building block-type compounds to the one-pot halogenation reaction (Fig. 3), selecting iodination for most substrates because of its convenient room temperature C–I bond-forming step. The process is compatible with pyridines with aryl, alkyl, and acetal substituents at the 2-position (**13** to **17**), and the acid-mediated conditions were effective for a series of pyridines (**18** to **21**). The one-pot iodination yield of 3-fluoropyridine (**20**) was low, and we reverted to a two-step process, isolating the intermediate Zincke imine. We recommend this alternative protocol to obtain higher overall yields of halogenated pyridines in these instances.

Pyridines with benzyl, cyano, and ketone substituents at the 4-position are also effective substrates (**22** to **25**). We then halogenated a series of 2,3-, 2,4-, and 2,5-disubstituted pyridines (**26** to **34**). Pyridine **27** is notable because of the sensitive 2-chloro substituent; in this case, ring opening with dibenzylamine resulted in a complex mixture of products. However, the reaction was successful using *N*-benzylaniline, and we attribute this outcome to its attenuated nucleophilicity. Halopyridines **32** to **34** show that the Zincke imine intermediate overrides the ortho-, para-directing benzyloxy group and maintains 3-selectivity. We formed a set of di- and trihalogenated pyridines from 3,4-disubstituted systems (**35** to **37**), and **38** to **40** are preliminary examples showing that this process can halogenate pyrimidines. When using 5-nitroisoquinoline as a substrate, we did not observe ring opening but instead isolated a dearomatized adduct where dibenzylamine adds to the C1 atom of the *N*Tf-isoquinolinium salt. Subjecting this intermediate to NBS then TFA resulted in 4-brominated product **41**. We observed a similar intermediate in a one-pot process using 5-azaindole, resulting in bromide **42**. A C6-brominated product **43** formed from 4-ethoxyquinoline; it is less clear what intermediate species operates in this case because of broadening in the ^{1}H NMR spectra after ring opening. Figure S2 shows proposed halogenation mechanisms and spectroscopic evidence of the dearomatized adducts.

Next, we tested whether halogenating Zincke imine intermediates could outcompete neighboring electron-rich aromatic rings and if dihalogenation processes were feasible. As shown in Fig. 3, a pyridine is selectively iodinated over an anisole group (**44**), as evidenced by only traces of other compounds, aside from starting material, in the crude ^{1}H NMR. As shown in fig. S10, we applied standard arene iodination conditions (NIS and TFA in MeCN at 50°C) and observed iodination on the anisole ring. Similarly, pyridines are iodinated over thiophenes, furans, and phenoxy groups (**45** to **48**). Furthermore, iodination selectively occurs on the pyridine in fused ring systems such as furopyridines, protected azaindoles, and thienopyridines that conventionally react at the five-membered heterocyclic portion (**49** to **51**) (*36*). C–Hal bond formation can also be applied sequentially to form 3,5-dihalogenated pyridines, a motif with two orthogonal vectors for further synthetic transformations (*37–39*). Although a one-pot procedure was unsuccessful, we were able to isolate Zincke imines **7-I** and **7-Br** and subject them to a second round of halogenation, resulting in various permutations of dihalopyridines **52** to **57**. We formed dichloropyridine **58** directly from Zincke imine **6** by modifying the acid-mediated halogenation-rearomatization protocol to include 2.05 equivalents of NCS.

The final stage of this study focused on halogenating complex drug-like intermediates, pharmaceuticals, and agrochemicals (Fig. 4). These structures often contain multiple (hetero)arenes, sites of reactivity, and Lewis basic atoms, thus representing a significant synthetic challenge for the ring-opening, halogenation, ring-closing process. However, success in this endeavor would enable subsequent convergent coupling reactions, compound diversification, and late-stage functionalization efforts that are not possible with existing pyridine halogenation methods (*40–42*). We began our study by halogenating five polycyclic structures representative of drug-like intermediates and obtained reasonable yields of halogenated products as single regioisomers in these cases (**59** to **65**). A substituted quinoline containing a cyclic sultam was again brominated at C6 (**66**). Halopyridines **67** to **70** are notable because C–Hal bond formation occurs on one pyridine over another; we believe that this selectivity results from exclusive *N*Tf-pyridinium salt formation involving the less hindered heterocycle in each case (*43*). Pyridines **71** and **72** formed atropisomers during the one-pot halogenation process, raising the possibility of controlling axial chirality in further studies.

This protocol successfully halogenated a diverse collection of pharmaceuticals and agrochemicals, demonstrating its utility for late-stage functionalization. We obtained moderate yields of iodinated vismodegib, bisacodyl, and nicoboxil (**73** to **75**). Loratadine proved challenging in the one-pot process, but we could synthesize iodo-, bromo-, and chloro-analogs when we isolated the intermediate Zincke imine (**76** to **78**). Nicotine, Boc-protected niaprazine, etoricoxib, and quinoxyfen were competent substrates (**79** to **83**). The dihalogenation sequence is also amenable to complex pyridines. Using the basal cell carcinoma drug vismodegib, we obtained iodo-chloro analog **84** in reasonable yield over two steps.

Given the lack of alternative methods for this transformation and the high value of the halopyridine products, we anticipate that this strategy will be useful for developing medicines and agrochemicals. This approach for pyridine functionalization is likely amenable to numerous other bond constructions as well.

REFERENCES AND NOTES

1. E. Vitaku, D. T. Smith, J. T. Njardarson, *J. Med. Chem.* **57**, 10257–10274 (2014).
2. M. Baumann, I. R. Baxendale, *Beilstein J. Org. Chem.* **9**, 2265–2319 (2013).
3. V. V. Zakharychev, A. V. Kuzenkov, A. M. Martsynkevich, *Chem. Heterocycl. Compd.* **56**, 1491–1516 (2020).
4. M. Stolar, T. Baumgartner, *Chem. Commun.* **54**, 3311–3322 (2018).
5. J. A. Joule, K. Mills, *Heterocyclic Chemistry* (Wiley, ed. 5, 2009).
6. M. R. Grimmett, *Adv. Heterocycl. Chem.* **58**, 271–345 (1993).
7. K. Murakami, S. Yamada, T. Kaneda, K. Itami, *Chem. Rev.* **117**, 9302–9332 (2017).
8. M. L. Crawley, B. M. Trost, *Applications of Transition Metal Catalysis in Drug Discovery and Development: An Industrial Perspective* (Wiley, 2012).
9. J. F. Bunnett, R. E. Zahler, *Chem. Rev.* **49**, 273–412 (1951).
10. W. F. Bailey, J. J. Patricia, *J. Organomet. Chem.* **352**, 1–46 (1988).
11. A. W. Hoffmann, *Ber. Dtsch. Chem. Ges.* **12**, 984–990 (1879).
12. W. J. Sell, F. W. Dootson, *J. Chem. Soc. Trans.* **73**, 442–445 (1898).
13. W. J. Sell, F. W. Dootson, *J. Chem. Soc. Trans.* **73**, 432–441 (1898).
14. P. S. Fier, J. F. Hartwig, *Science* **342**, 956–960 (2013).
15. G. A. Olah, *Acc. Chem. Res.* **4**, 240–248 (1971).
16. B. Galabov, D. Nalbantova, Pv. Schleyer, H. F. Schaefer3rd, *Acc. Chem. Res.* **49**, 1191–1199 (2016).
17. S. M. Manolikakes, N. M. Barl, C. Samann, P. Knochel, *Z. Naturforsch. B. J. Chem. Sci.* **68**, 411–422 (2013).
18. G. A. El-Hiti, K. Smith, A. S. Hegazy, M. B. Alshammari, A. M. Masmali, *ChemInform* **46**, 1 (2015).
19. G. A. El-Hiti, K. Smith, A. S. Hegazy, *Heterocycles* **91**, 479–504 (2015).
20. M. A. Larsen, J. F. Hartwig, *J. Am. Chem. Soc.* **136**, 4287–4299 (2014).
21. S. A. Sadler *et al.*, *Org. Biomol. Chem.* **12**, 7318–7327 (2014).
22. J. S. Wright, P. J. H. Scott, P. G. Steel, *Angew. Chem. Int. Ed.* **60**, 2796–2821 (2021).
23. L. Yang, N. Uemura, Y. Nakao, *J. Am. Chem. Soc.* **141**, 7972–7979 (2019).
24. J. Trouvé, P. Zardi, S. Al-Shehimy, T. Roisnel, R. Gramage-Doria, *Angew. Chem. Int. Ed.* **60**, 18006–18013 (2021).
25. C. Cheng, J. F. Hartwig, *J. Am. Chem. Soc.* **137**, 592–595 (2015).
26. X. Y. Zhou, M. Zhang, Z. Liu, J. H. He, X. C. Wang, *J. Am. Chem. Soc.* **144**, 14463–14470 (2022).
27. C. D. Vanderwal, *J. Org. Chem.* **76**, 9555–9567 (2011).
28. W. C. Cheng, M. J. Kurth, *Org. Prep. Proced. Int.* **34**, 585–608 (2002).
29. S. Sowmiah, J. M. S. S. Esperanca, L. P. N. Rebelo, C. A. M. Afonso, *Org. Chem. Front.* **5**, 453–493 (2018).
30. R. A. Toscano *et al.*, *Chem. Pharm. Bull. (Tokyo)* **45**, 957–961 (1997).
31. J. Becher, *Synthesis* **1980**, 589–612 (1980).

32. U. Stämpfli, M. Neuenschwander, *Helv. Chim. Acta* **66**, 1427–1435 (1983).
33. T. M. Nguyen, M. del Rayo Sanchez-Salvatori, J. C. Wypych, C. Marazano, *J. Org. Chem.* **72**, 5916–5919 (2007).
34. A. Tomberg, M. J. Johansson, P. O. Norrby, *J. Org. Chem.* **84**, 4695–4703 (2019).
35. M. Liljenberg, J. H. Stenlid, T. Brinck, *J. Phys. Chem. A* **122**, 3270–3279 (2018).
36. L. H. Klemm, J. N. Louris, *J. Heterocycl. Chem.* **21**, 785–789 (1984).
37. E. K. Reeves, E. D. Entz, S. R. Neufeldt, *Chemistry* **27**, 6161–6177 (2021).
38. M. H. Keylor, Z. L. Niemeyer, M. S. Sigman, K. L. Tan, *J. Am. Chem. Soc.* **139**, 10613–10616 (2017).
39. J. Ji, T. Li, W. H. Bunnelle, *Org. Lett.* **5**, 4611–4614 (2003).
40. T. Cernak, K. D. Dykstra, S. Tyagarajan, P. Vachal, S. W. Krska, *Chem. Soc. Rev.* **45**, 546–576 (2016).
41. L. Zhang, T. Ritter, *J. Am. Chem. Soc.* **144**, 2399–2414 (2022).
42. D. C. Blakemore *et al*., *Nat. Chem.* **10**, 383–394 (2018).
43. R. D. Dolewski, P. J. Fricke, A. McNally, *J. Am. Chem. Soc.* **140**, 8020–8026 (2018).

ACKNOWLEDGMENTS

Funding: This work was supported by Colorado State University with funds from the National Institutes of Health (NIH grant R01 GM144591-01). R.S.P. acknowledges support from the National Science Foundation (NSF grant CHE-1955876); use of computational resources from the RMACC Summit Supercomputer, which is funded by NSF grants ACI-1532235 and ACI-1532236, the University of Colorado Boulder, and Colorado State University; and support from XSEDE through allocation TG-CHE180056. **Author contributions:** B.T.B. and J.N.L. performed the experimental work. A.M., B.T.B., and J.N.L. conceptualized the work. R.S.P. and L.d.L. performed the computational studies. All authors contributed to the design of the experimental and computational work, performed data analysis, discussed the results, and commented on the manuscript. A.M. and R.S.P. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. Cartesian coordinates and absolute thermochemical data for all stationary points in computational studies are available in data S1.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add8980
Materials and Methods
Figs. S1 to S23
Tables S1 to S12
NMR Spectra
References (*44*–*74*)
Data S1

Submitted 12 July 2022; accepted 7 October 2022
10.1126/science.add8980

ORGANIC CHEMISTRY

Radical and ionic *meta*-C–H functionalization of pyridines, quinolines, and isoquinolines

Hui Cao†, Qiang Cheng†, Armido Studer*

Carbon-hydrogen (C–H) functionalization of pyridines is a powerful tool for the rapid construction and derivatization of many agrochemicals, pharmaceuticals, and materials. Because of the inherent electronic properties of pyridines, selective *meta*-C–H functionalization is challenging. Here, we present a protocol for highly regioselective *meta*-C–H trifluoromethylation, perfluoroalkylation, chlorination, bromination, iodination, nitration, sulfanylation, and selenylation of pyridines through a redox-neutral dearomatization-rearomatization process. The introduced dearomative activation mode provides a diversification platform for meta-selective reactions on pyridines and other azaarenes through radical as well as ionic pathways. The broad scope and high selectivity of these catalyst-free reactions render these processes applicable for late-stage functionalization of drugs.

Pyridines and their derivatives are among the most frequently occurring heteroarenes in pharmaceuticals and agrochemicals that contain N-heterocycles (Fig. 1A) (*1*, *2*). They also play prominent roles in ligands as well as functional materials (*3*, *4*). Selective modification of pyridines without preinstalled transformable functional groups can substantially increase step economy for the preparation of related, more elaborate compounds (*5*). The known methods for carbon-hydrogen (C–H) functionalization of pyridines mainly rely on the electronically biased reactivity of the ortho- and para-positions owing to the electronic deficiency of the π-system as well as the efficient σ-donor property of the nitrogen atom (*5*, *6*). Current strategies for direct functionalization of pyridines include directed metalation (*7*, *8*), Minisci-type radical reactions (*9*), and nucleophilic addition to N-activated pyridines (*10*, *11*). However, for unbiased pyridines, direct meta-functionalization is far more challenging, with limited documented examples (*12*). In particular, meta-selective processes proceeding in both radical and ionic reactivity modes are unavailable.

Although electrophilic aromatic substitutions have been applied to meta-selective halogenation and nitration of pyridines, the harsh acidic conditions limit their usage and frequently lead to regioisomeric products (*13*, *14*). Milder meta-functionalization reactions have been developed in recent decades, mainly through transition-metal catalysis (Fig. 1B) (*15*). On the basis of steric control, iridium-catalyzed C–H borylation and silylation reactions occur regioselectively at the meta position (*16*–*21*). Palladium-catalyzed *meta*-C–H olefination and arylation reactions were pioneered by the Yu group by using designed ligands for regioselectivity control (*22*–*25*). Despite the extensive development and application of those transition metal–catalyzed reactions, the reaction type and scope are still limited; for example, 3-substituted pyridines are required for regioselectivity control in most cases (*15*).

An alternative strategy for the meta-functionalization of pyridines is the temporary dearomatization approach, so that the initially electron-deficient heteroarenes are transformed into activated electron-rich intermediates. The ensuing electrophilic reactions and subsequent rearomatization provide exclusively meta-substituted heteroarenes (Fig. 1C) (*26*–*31*). Several strategies for the reductive dearomatization of pyridines have been developed along those lines (*32*, *33*). However, most of them generate either stable enamines so that rearomatization becomes energetically unfavorable, or alternatively very unstable intermediate *N*-silyl or *N*-boryl species that constrain applicable functionalization reactions. For example, the interrupted reductive dearomatization of pyridiniums allowed for subsequent selective C3-functionalization, but tetrahydropyridines instead of the aromatized heteroarenes were formed as the final products (*34*). meta-Selective silylation, alkylation, and trifluoromethylthiolation have been developed through 1,4-reduction of pyridines by silanes or boranes followed by in situ functionalization and oxidative rearomatization (*28*–*30*). Nevertheless, introduction of the trifluoromethyl and halogen groups to these reductively dearomatized intermediates remains unachieved, most likely because of the sensitivity of these intermediates to oxidation in the presence of electrophilic reagents (*35*). To date, none of the reported methods provide access to stable dearomatized intermediates of pyridines that possess enough redox stability to undergo radical addition and electrophilic halogenation instead of direct oxidative rearomatization yet still rearomatize under specific conditions, such as acid treatment. We present a protocol for versatile and practical *meta*-C–H functionalization of pyridines, quinolines, and isoquinolines through a redox-neutral dearomatization-rearomatization process (Fig. 1D). Pyridines can react with inexpensive and commercially available acetylenedicarboxylates to form Huisgen 1,4-dipoles, which readily undergo high-yielding dearomative cycloaddition reactions with carbonyl dipolarophiles (*36*, *37*). We found that the formed oxazino pyridines are bench-stable intermediates, which could engage

Organisch-Chemisches Institut, Westfälische Wilhelms-Universität, Münster, Germany.
*Corresponding author. Email: studer@uni-muenster.de
†These authors contributed equally to this work.

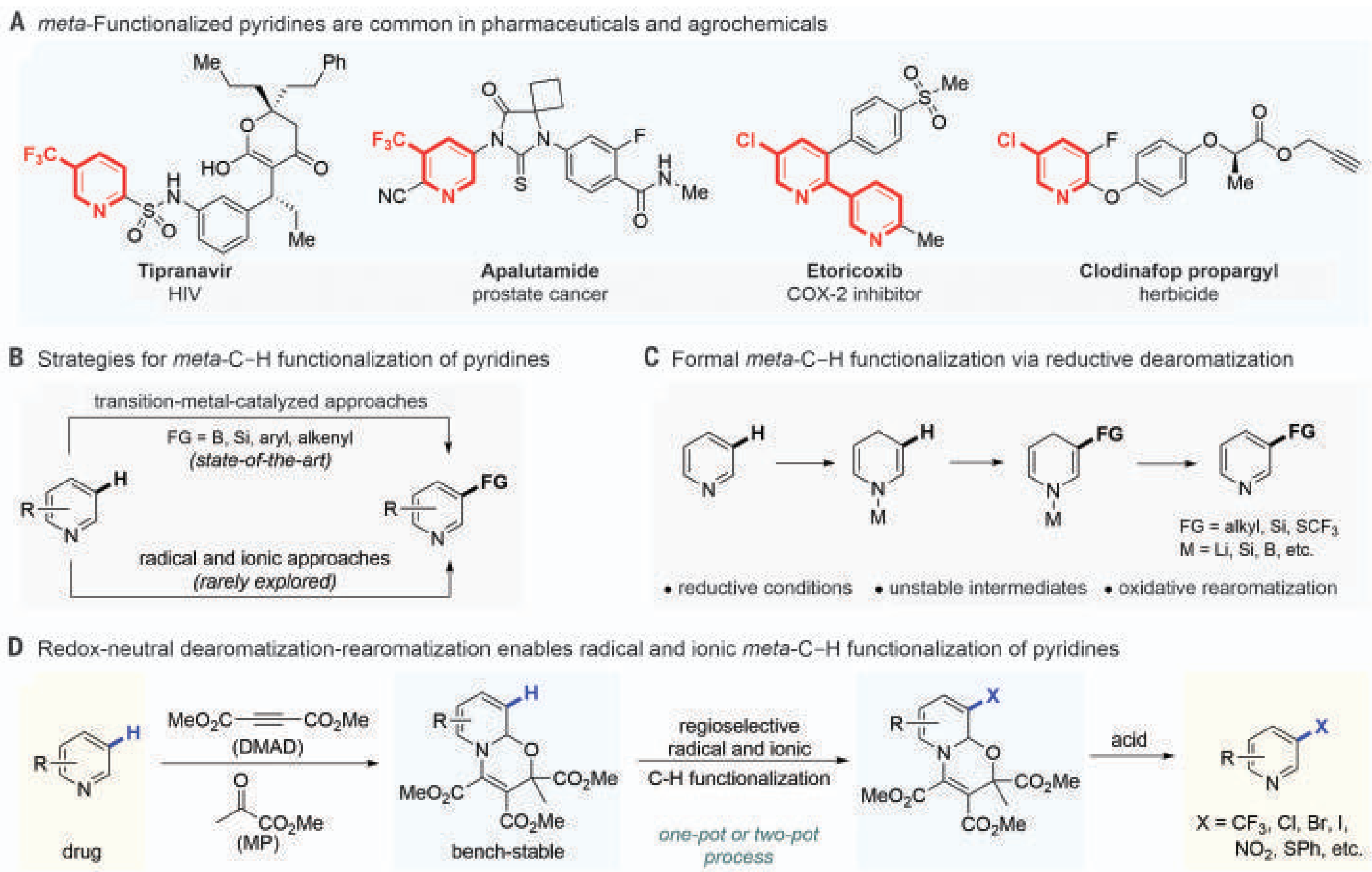


Fig. 1. meta-Functionalization of pyridines. (**A**) Various biologically active compounds containing a meta-functionalized pyridine moiety. (**B**) Established routes for the direct *meta*-C–H functionalization of pyridines. (**C**) meta-Functionalization through reduction to the dihydropyridine. (**D**) Developed strategy through redox-neutral activation of pyridines and subsequent regioselective ionic and radical C–H functionalization. FG, functional group; DMAD, dimethyl acetylenedicarboxylate; MP, methyl pyruvate.

in regioselective radical and ionic functionalizations by use of a variety of radical precursors and electrophilic reagents. Acid-promoted rearomatization then provides the corresponding meta-functionalized pyridines. With the redox-neutral activation method of N-heteroarenes, we could achieve exclusively meta-selective C–H functionalization to form trifluoromethyl, perfluoroalkyl, chloro, bromo, iodo, nitro, thio, and seleno N-heteroarenes under mild conditions.

We evaluated first the reaction conditions for redox-neutral dearomatization and rearomatization of pyridines (Fig. 2). A broad range of pyridines engaged in three-component coupling with dimethyl acetylenedicarboxylate (DMAD) and methyl pyruvate (MP) in acetonitrile at room temperature under air (tables S1 to S3), with an excellent average yield of 89% (fig. S5). The generated oxazino pyridine intermediates, used as mixtures of diastereoisomers [in some cases also as a regioisomeric mixture (fig. S5)], were stable toward air, water, and silica-gel-column chromatography (fig. S7), and these activated pyridines could then be used as efficient carbon-radical acceptors as well as nucleophiles in reactions with various electrophilic trapping reagents. Separation of the isomers was not required because they converged to the same product after functionalization and rearomatization. We found that rearomatization of the regioselectively C–H functionalized oxazino pyridine intermediates is readily achieved in aqueous acid at 60°C in quantitative yields (table S4).

Considering the radical *meta*-pyridine functionalization, C–H trifluoromethylation of the oxazino pyridine intermediates merits attention. Although radical trifluoromethylation is a practical method for installing the pharmaceutically highly relevant trifluoromethyl group on arenes (*38*), poor regioselectivity has been observed for unbiased pyridines owing to polarity-mismatch (*39*). A general strategy to directly install a trifluoromethyl or other perfluoroalkyl groups selectively at the meta-position of pyridines under mild conditions is highly desirable. In this work, we realized formal direct meta-selective trifluoromethylation of pyridines by merging the dearomatization-rearomatization strategy with trifluoromethyl radical chemistry. Trifluoroiodomethane, one of the cheapest trifluoromethyl sources (*38*), together with the organic base DBU (1,8-diazabicyclo[5.4.0]undec-7-ene) effectively promote trifluoromethylation of the oxazino pyridine intermediates under blue light irradiation. We propose that the generated electrophilic trifluoromethyl radical (figs. S12 to S14) should electronically match with the nucleophilic dearomatized intermediates (*40*), which should also be responsible for the high regioselectivity. Blue light was required for chain initiation through carbon-iodine (C–I) bond homolysis, and we assumed that the radical generated by the addition of the trifluoromethyl radical to the oxazino pyridine intermediate further reacts through iodine atom abstraction from CF_3I, sustaining the chain (fig. S14). DBU-mediated HI elimination leads to the C–H trifluoromethylated oxazino pyridine, driving the endothermic I atom transfer step (*41*).

Both electron-donating phenyl (**2**), phenoxy (**3**), alkyl groups (**7**), and electron-withdrawing halo substituents (**9**, **10**, **11**, and **16**) on pyridines at different positions were compatible with the reaction. For substrates with two *meta*-C–H positions available (**1** to **6**), including unsubstituted pyridine (**1**), only mono-functionalized

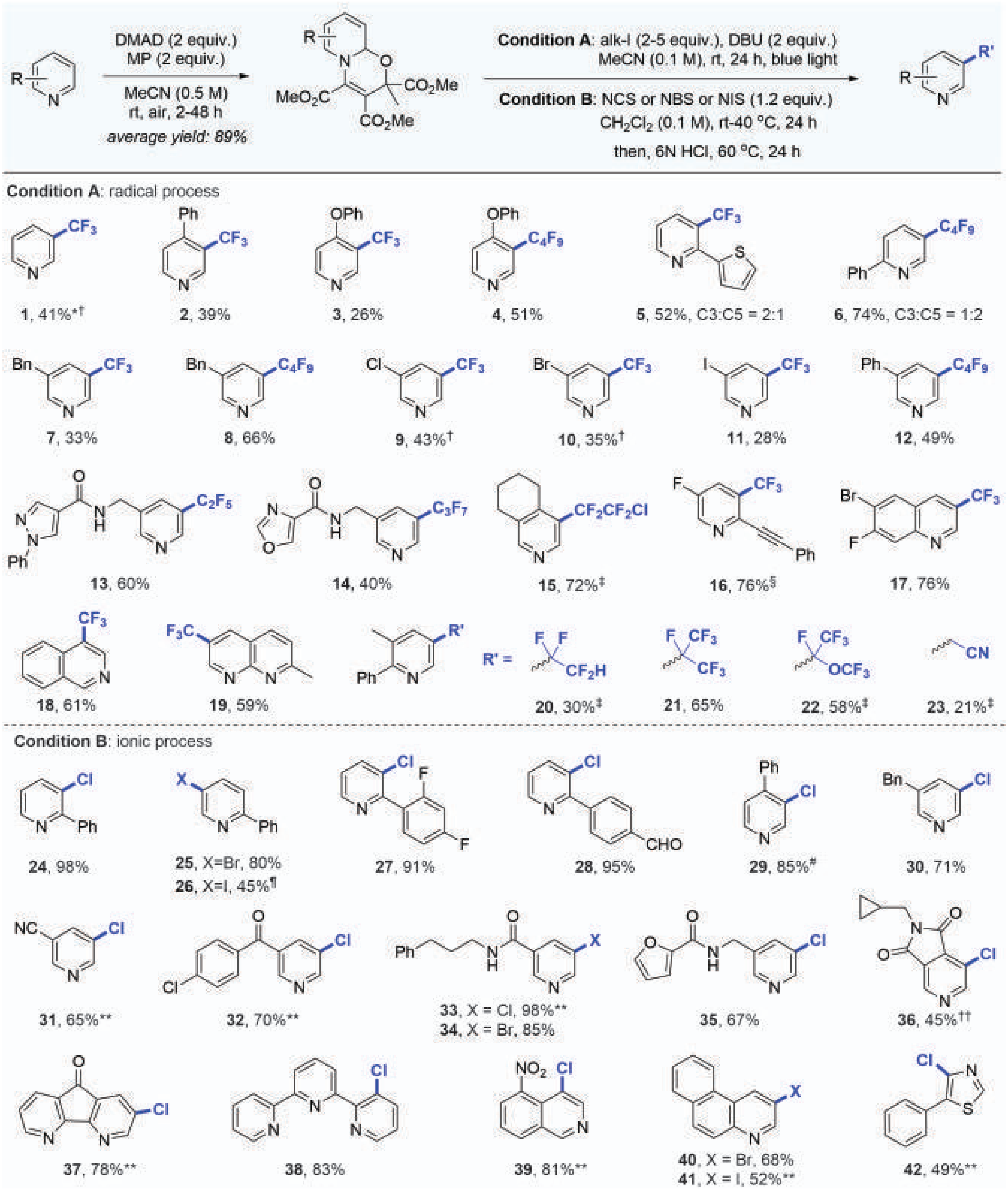


Fig. 2. Scope of radical meta-fluoroalkylation and ionic meta-halogenation. Compounds **1** to **23** are accessible through radical C–H functionalization by using condition A. Ionic functionalization leads to halogenated heteroarenes **24** to **42** by applying condition B. All yields of pyridine products correspond to isolated yields on the basis of the isolated dearomatized intermediates. Asterisk indicates 1.3 equiv alkyl iodide. Dagger symbol (†) indicates that yield was determined by the analysis of ^1H nuclear magnetic resonance (NMR) spectra of crude product by using dibromomethane as an internal standard. Double-dagger symbol (‡) indicates 10 equiv alkyl iodide. Section symbol (§) indicates that H_2SO_4 was used instead of HCl. Paragraph symbol (¶) indicates 60% conversion based on the analysis of ^1H NMR spectra of crude product by using dibromomethane as an internal standard. Pound symbol (#) indicates mono- and di-chlorination at 3- and 5-positions, with a ratio of 5:1 (mono:di). Two asterisks (**) indicate trimethyl silyl chloride (2.0 equiv) as additive, heating at 40°C. Two dagger symbols (††) indicate trimethyl silyl chloride (2.0 equiv) as additive, heating at 60°C. DMAD, dimethyl acetylenedicarboxylate; MP, methyl pyruvate; MeCN, acetonitrile; DBU, 1,8-diazabicyclo[5.4.0]undec-7-ene; NCS, *N*-chlorosuccinimide; NBS, *N*-bromosuccinimide; NIS, *N*-iodosuccinimide.

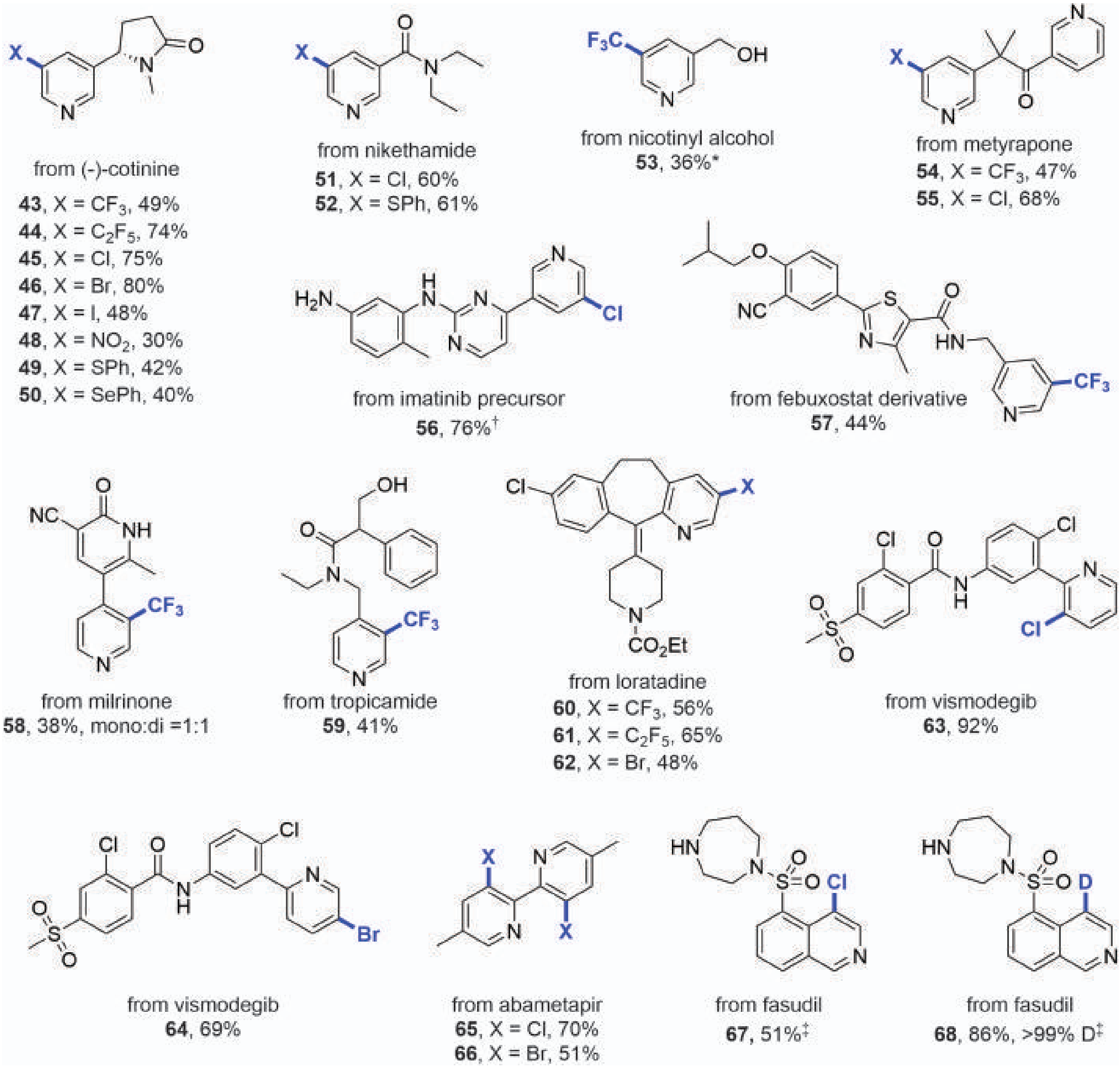


Fig. 3. meta-Functionalization of drugs and drug derivatives. Radical trifluoromethylation of various heteroarene containing biologically active molecules is achieved. The ionic C–H functionalization can be used for *meta*-C–H halogenation, nitration, sulfanylation, selenylation, and deuteration on more complex heteroarenes. All yields of pyridine products correspond to isolated yields on the basis of the isolated dearomatized intermediates. Experimental details are provided in the supplementary materials. Asterisk indicates from the corresponding acetate. Dagger symbol (†) indicates from the corresponding acetamide. Double-dagger symbol (‡) indicates from the corresponding *tert*-butyl carbamate.

products were obtained. The dienamine entity of the oxazino pyridine intermediates was reactive at the β- and δ-position, and 2-aryl pyridines gave a mixture of 3- and 5-functionalized products (**5** and **6**). The reverse selectivity for products **5** and **6** indicates the inherent steric bias of the radical addition process. Chloro-, bromo-, and iodo-trifluoromethylated pyridines (**9**, **10**, and **11**, respectively), which can serve as versatile precursors to drug-like compounds, were accessible in moderate yields. The reaction was also compatible with other heteroarenes, such as thiophenes (**5**), pyrazoles (**13**), and oxazoles (**14**), with activation (dearomatization) and accordingly C–H functionalization chemoselectively occurring on the pyridine moiety. In general, electron-rich arenes such as thiophenes are better trifluoromethyl radical acceptors than pyridines (*39*), showing that our activation mode overrides innate reactivity patterns. Along these lines, our strategy enables selective trifluoromethylation of pyridines in the presence of electron-rich arenes (**3** and **5**). Multisubstituted pyridines (**15**, **16**, and **20** to **23**) were compatible substrates. Quinolines (**17**), isoquinolines (**18**), and naphthyridines (**19**) also engaged in the transformation on application of the same activation strategy to provide the corresponding trifluoromethylated heteroarenes with

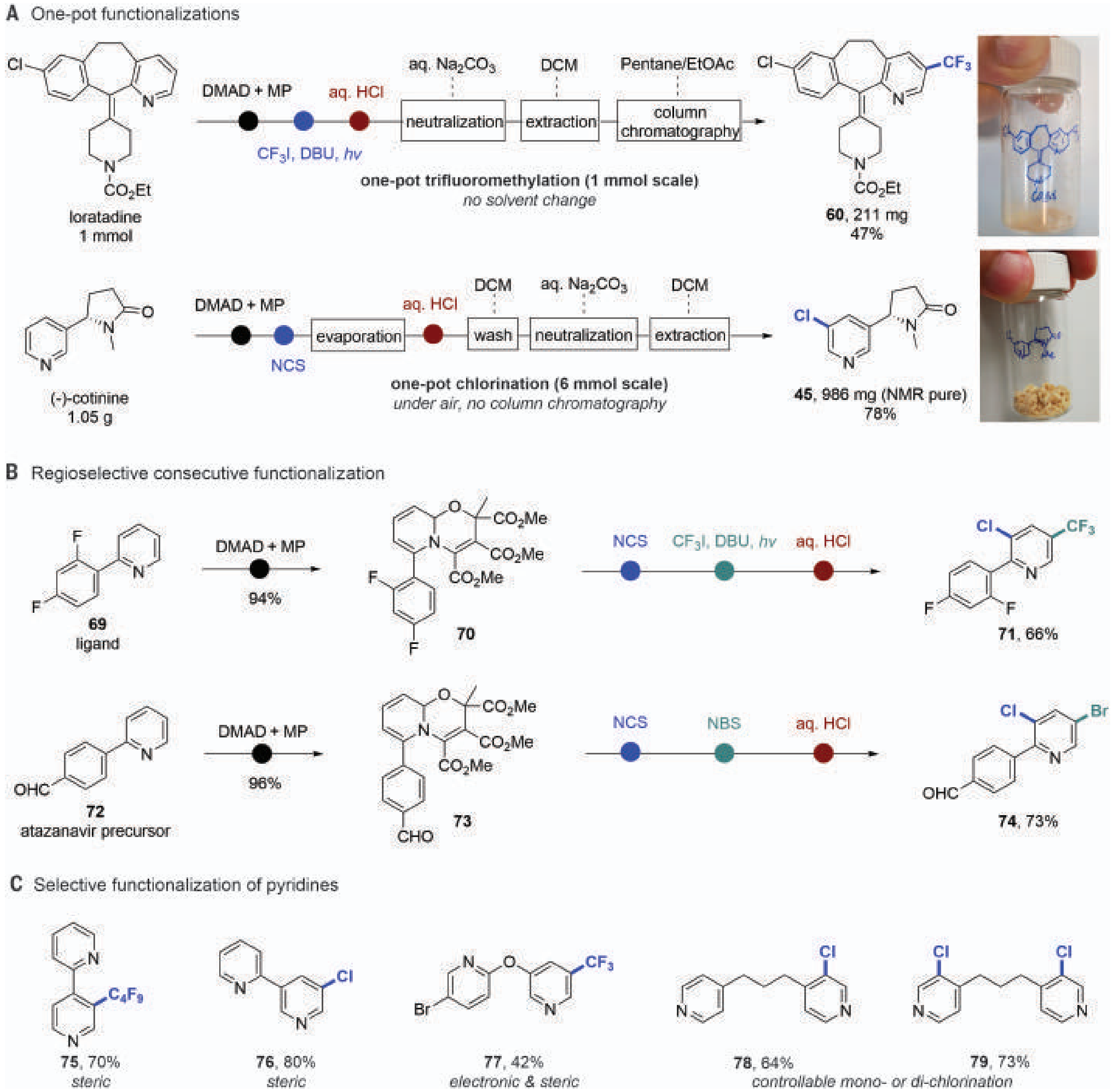


Fig. 4. Synthetic applications. (**A**) meta-Trifluoromethylation of loratadine and meta-chlorination of (−)-cotinine as one-pot processes on a larger scale. (**B**) Sequential radical-ionic or ionic-ionic meta,meta′-difunctionalization of pyridines. (**C**) Chemoselective *meta*-C–H functionalization of molecules containing two pyridine rings.

complete meta-selectivity. Along with the trifluoromethyl group, other polyfluoroalkyl groups (**4**, **6**, **8**, **12** to **15**, and **20** to **22**) could also be introduced to the meta position of pyridines, quinolines, and isoquinolines under the same conditions by simply varying the radical precursors. As compared with the trifluoromethylation, perfluoroalkylation with the corresponding commercial iodides worked with higher efficiency. The relatively low yields of the trifluoromethylation reactions were mainly due to low conversions (figs. S10 and S11). Varying the oxazino pyridine moiety by using different dipolarophiles for pyridine dearomatization did not provide better results (figs. S8 and S9). With iodoacetonitrile, meta-alkylation of pyridines could be achieved, albeit in low yield (**23**).

Direct meta-halogenation of pyridines under mild conditions has been a longstanding challenge in synthetic chemistry (*42*). Using the redox-neutral dearomatization activation strategy, selective electrophilic halogenation at the meta-position of pyridines possessing both electron-rich and electron-poor substituents was readily achieved with commercially available *N*-halosuccinimides (**24** to **42**). For 2-aryl pyridines with two meta-positions available, mono-chlorination and mono-bromination were achieved with opposite regioselectivity (**24** to **28**) as chlorination occurred exclusively at the 3-position, whereas bromination proceeded at the 5-position, addressing the

β- or δ-reactivity of the dienamine intermediate, respectively. A plausible explanation is that chlorination occurs irreversibly at the more nucleophilic β-position, delivering the kinetic product, whereas because of the weaker carbon-bromine (C–Br) bond, bromination is reversible and leads to the less sterically hindered and thermodynamically more stable δ-bromo dienamine intermediates (*42*, *43*). This analysis is in line with the iodination that also occurs exclusively at the 5-position (**26**). We observed good selectivity of mono-chlorination for 4-aryl pyridines (**29**). Activation of the chlorinating reagent, *N*-chlorosuccinimide, with trimethylsilyl chloride (*44*) further extended the scope to more electron-deficient pyridines (**31** to **34**, **36**, and **37**) that are virtually unreactive under conventional electrophilic aromatic halogenation conditions (*13*). Moreover, bipyridine and terpyridine ligands, such as **37** and **38**, could be selectively mono-dearomatized at one of the pyridine rings, delivering after rearomatization monochlorinated products. Isoquinolines (**39**) and 1-azaphenanthrenes (**40** and **41**) can also be selectively halogenated at the meta position of the pyridine core. The potential application of this redox-neutral activation strategy toward other N-heteroarenes could be demonstrated by the successful 4-selective chlorination of a thiazole compound (**42**).

Late-stage functionalization enables rapid modification of drugs, drug candidates, and materials without the need for de novo synthesis (*45*, *46*). We demonstrated that the developed redox-neutral dearomatization-functionalization-rearomatization strategy is a practical method for the direct late-stage modification of drugs that contain pyridine moieties (Fig. 3). Twelve drugs and drug derivatives with varied substitution patterns underwent diverse meta-selective functionalizations in generally good yields (30 to 92%; **43** to **68**). As an illustration, various functionalities—including trifluoromethyl (**43**), pentafluoroethyl (**44**), chloro (**45**), bromo (**46**), iodo (**47**), nitro (**48**), sulfanyl (**49**), and selenyl (**50**) groups—could be selectively introduced into the meta position of the alkaloid (–)-cotinine through radical and ionic pathways. Although some drugs and precursors required protection of the hydroxyl or amino groups before dearomatization (**53**, **56**, **67**, and **68**), acid-promoted rearomatization expediently furnished the free alcohol or amine products. The relatively electron-rich pyridine ring in metyrapone was selectively trifluoromethylated (**54**) and chlorinated (**55**) in 47 and 68% yield, respectively. The pyrimidine moiety in the imatinib precursor was well tolerated (**56**; 76%). Tropicamide with a 4-alkyl pyridine moiety was mono-trifluoromethylated in moderate yield (**59**), whereas milrinone with a 4-aryl pyridine moiety delivered an isolable mixture of mono- and di-trifluoromethylated products (**58**). Vismodegib furnished 3-chlorinated (**63**; 92%) and 5-brominated (**64**; 69%) products with exclusive regiocontrol. For the drug abametapir with a bipyridine structure, the mono-dearomatized intermediate gave meta, meta′-dichlorination (**65**; 70%) and meta,meta′-dibromination (**66**; 51%) products, probably because of high reactivity of the corresponding dearomatized intermediate. To illustrate the potential use of this redox-neutral activation method for other previously inaccessible meta-functionalization reactions, we achieved selective deuteration of fasudil (**68**; 86%, >99% deuterium) by treating the dearomatized intermediate with deuterated trifluoroacetic acid and deuterium oxide (D_2O).

The meta-functionalizations of pyridines described above were carried out by means of two-pot processes, in which the dearomatized intermediates were isolated before the ensuing C–H functionalization step to guarantee facile isolation of products and limit side reactions. This allowed for ready diversification because various radical precursors and electrophiles engage in the C–H functionalization, as we have documented. However, targeting a single meta-functionalized compound, we found that most reactions can be readily conducted as one-pot processes. To demonstrate the practicality and robustness of the method in both medicinal and process chemistry, we displayed one-pot reactions at a larger scale, starting from loratadine and (–)-cotinine (Fig. 4A and figs. S3 and S4). meta-Trifluoromethylation of loratadine on a 1-mmol scale gave 47% isolated yield of **60** without any solvent change. Gram-scale synthesis of meta-chlorinated (–)-cotinine **45** was achieved under air without the need for column chromatography. Through sequential ionic chlorination and radical trifluoromethylation, regioselective consecutive functionalization of a ligand was realized to generate 3,5-difunctionalized pyridine **71** (66% yield) without isolation of the mono-functionalized intermediates, leveraging the potential of the β- and δ-reactivity of the dienamine intermediates (Fig. 4B). Likewise, sequential ionic-ionic difunctionalization of the atazanavir precursor was accomplished (**74**; 74%). When two differently substituted pyridines were presented in the same molecule, the redox-neutral dearomatization showed high selectivity toward more electron-rich and less sterically hindered pyridines, allowing for chemoselective mono-meta-functionalization of polypyridine compounds [**54** and **55** (Fig. 3) and **75** to **79** (Fig. 4C)]. Even for compounds that contain pyridines in the same chemical environment, such as the linker molecule used to construct metal organic frameworks (**78** and **79**) (*47*), controlled mono- and di-dearomatization can lead to the corresponding mono- and di-chlorination products, respectively, in high yields.

We have developed a general method for radical and ionic *meta*-C–H functionalization of pyridines, quinolines, and isoquinolines through a redox-neutral dearomatization-rearomatization strategy. The high selectivity, robust reaction operations, broad scope of transformations, and late-stage application all contribute to the practicality of the method. Therefore, this redox-neutral activation approach for meta-selective functionalization of heteroarenes should be widely applicable in the pharmaceutical and agrochemical research arenas.

REFERENCES AND NOTES

1. M. Baumann, I. R. Baxendale, *Beilstein J. Org. Chem.* **9**, 2265–2319 (2013).
2. E. Vitaku, D. T. Smith, J. T. Njardarson, *J. Med. Chem.* **57**, 10257–10274 (2014).
3. J. K. Kallitsis, M. Geormezi, S. G. Neophytides, *Polym. Int.* **58**, 1226–1233 (2009).
4. M. N. Zafar *et al.*, *Russ. J. Coord. Chem.* **42**, 1–18 (2016).
5. K. Murakami, S. Yamada, T. Kaneda, K. Itami, *Chem. Rev.* **117**, 9302–9332 (2017).
6. F.-Y. Zhou, L. Jiao, *Synlett* **32**, 159–178 (2021).
7. M. Balkenhohl, P. Knochel, *SynOpen* **2**, 0078–0095 (2018).
8. Y. Nakao, *Synthesis* **2011**, 3209–3219 (2011).
9. R. S. J. Proctor, R. J. Phipps, *Angew. Chem. Int. Ed.* **58**, 13666–13699 (2019).
10. J. A. Bull, J. J. Mousseau, G. Pelletier, A. B. Charette, *Chem. Rev.* **112**, 2642–2713 (2012).
11. R. D. Dolewski, M. C. Hilton, A. McNally, *Synlett* **29**, 08–14 (2018).
12. D. E. Stephens, O. V. Larionov, *Tetrahedron* **71**, 8683–8716 (2015).
13. G. A. Olah, *Acc. Chem. Res.* **4**, 240–248 (1971).
14. B. Galabov, D. Nalbantova, Pv. Schleyer, H. F. Schaefer 3rd, *Acc. Chem. Res.* **49**, 1191–1199 (2016).
15. R. Das, M. Kapur, *Asian J. Org. Chem.* **7**, 1217–1235 (2018).
16. M. A. Larsen, J. F. Hartwig, *J. Am. Chem. Soc.* **136**, 4287–4299 (2014).
17. J. S. Wright, P. J. H. Scott, P. G. Steel, *Angew. Chem. Int. Ed.* **60**, 2796–2821 (2021).
18. S. A. Sadler *et al.*, *Org. Biomol. Chem.* **12**, 7318–7327 (2014).
19. L. Yang, N. Uemura, Y. Nakao, *J. Am. Chem. Soc.* **141**, 7972–7979 (2019).
20. C. Cheng, J. F. Hartwig, *J. Am. Chem. Soc.* **137**, 592–595 (2015).
21. B.-J. Li, Z.-J. Shi, *Chem. Sci.* **2**, 488–493 (2011).
22. M. Ye, G.-L. Gao, J.-Q. Yu, *J. Am. Chem. Soc.* **133**, 6964–6967 (2011).
23. M. Ye *et al.*, *J. Am. Chem. Soc.* **133**, 19090–19093 (2011).
24. T. Zhang *et al.*, *Nat. Chem.* **13**, 1207–1213 (2021).
25. Z. Fan *et al.*, *Nature* **610**, 87–93 (2022).
26. C. S. Giam, S. D. Abbott, *J. Am. Chem. Soc.* **93**, 1294–1296 (1971).
27. O. Tsuge, S. Kanemasa, T. Naritomi, J. Tanaka, *Chem. Lett.* **13**, 1255–1258 (1984).
28. S. Wübbolt, M. Oestreich, *Angew. Chem. Int. Ed.* **54**, 15876–15879 (2015).
29. Z. Liu *et al.*, *J. Am. Chem. Soc.* **144**, 4810–4818 (2022).
30. X.-Y. Zhou, M. Zhang, Z. Liu, J.-H. He, X.-C. Wang, *J. Am. Chem. Soc.* **144**, 14463–14470 (2022).
31. A. Grozavu, H. B. Hepburn, E. P. Bailey, P. J. Lindsay-Scott, T. J. Donohoe, *Chem. Sci.* **11**, 8595–8599 (2020).
32. S. Park, S. Chang, *Angew. Chem. Int. Ed.* **56**, 7720–7738 (2017).
33. G. Bertuzzi, L. Bernardi, M. Fochi, *Catalysts* **8**, 632 (2018).
34. A. Grozavu *et al.*, *Nat. Chem.* **11**, 242–247 (2019).
35. M. W. Gribble Jr., S. Guo, S. L. Buchwald, *J. Am. Chem. Soc.* **140**, 5057–5060 (2018).
36. R. Huisgen, M. Morikawa, K. Herbig, E. Brunn, *Chem. Ber.* **100**, 1094–1106 (1967).
37. V. Nair, A. Deepthi, D. Ashok, A. E. Raveendran, R. R. Paul, *Tetrahedron* **70**, 3085–3105 (2014).
38. H. Baguia, G. Evano, *Chem. Eur. J.* **28**, e202200975 (2022).
39. F. O'Hara, D. G. Blackmond, P. S. Baran, *J. Am. Chem. Soc.* **135**, 12122–12134 (2013).
40. A. Studer, *Angew. Chem. Int. Ed.* **51**, 8950–8958 (2012).
41. F. Juliá, T. Constantin, D. Leonori, *Chem. Rev.* **122**, 2292–2352 (2022).
42. B. T. Boyle, J. N. Levy, L. de Lescure, R. S. Paton, A. McNally, ChemRxiv 88802 [Preprint] (2022); doi:10.26434/chemrxiv-2022-88802-v3.
43. L. A. Paquette, W. C. Farley, *J. Org. Chem.* **32**, 2725–2731 (1967).

44. T. Maibunkaew, C. Thongsornkleeb, J. Tummatorn, A. Bunrit, S. Ruchirawat, *Synlett* **25**, 1769–1775 (2014).
45. T. Cernak, K. D. Dykstra, S. Tyagarajan, P. Vachal, S. W. Krska, *Chem. Soc. Rev.* **45**, 546–576 (2016).
46. J. Börgel, T. Ritter, *Chem* **6**, 1877–1887 (2020).
47. F. A. Almeida Paz, J. Klinowski, *Inorg. Chem.* **43**, 3882–3893 (2004).

ACKNOWLEDGMENTS

Funding: We thank the Westfälische Wilhelms–Universität Münster for supporting this work. **Author contributions:** H.C., Q.C., and A.S. designed and analyzed the experiments. H.C. and Q.C. conducted the experiments. H.C., Q.C., and A.S. wrote the manuscript. A.S. directed the project. All authors contributed to discussions. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade6029
Materials and Methods
Supplementary Text
Figs. S1 to S14
Tables S1 to S4
NMR Spectra
References (*48–56*)

Submitted 26 August 2022; accepted 18 October 2022
10.1126/science.ade6029

QUANTUM PHYSICS

Noise-resilient edge modes on a chain of superconducting qubits

X. Mi[1], M. Sonner[2], M. Y. Niu[1], K. W. Lee[1], B. Foxen[1], R. Acharya[1], I. Aleiner[1], T. I. Andersen[1], F. Arute[1], K. Arya[1], A. Asfaw[1], J. Atalaya[1], J. C. Bardin[1,3], J. Basso[1], A. Bengtsson[1], G. Bortoli[1], A. Bourassa[1], L. Brill[1], M. Broughton[1], B. B. Buckley[1], D. A. Buell[1], B. Burkett[1], N. Bushnell[1], Z. Chen[1], B. Chiaro[1], R. Collins[1], P. Conner[1], W. Courtney[1], A. L. Crook[1], D. M. Debroy[1], S. Demura[1], A. Dunsworth[1], D. Eppens[1], C. Erickson[1], L. Faoro[1], E. Farhi[1], R. Fatemi[1], L. Flores[1], E. Forati[1], A. G. Fowler[1], W. Giang[1], C. Gidney[1], D. Gilboa[1], M. Giustina[1], A. G. Dau[1], J. A. Gross[1], S. Habegger[1], M. P. Harrigan[1], M. Hoffmann[1], S. Hong[1], T. Huang[1], A. Huff[1], W. J. Huggins[1], L. B. Ioffe[1], S. V. Isakov[1], J. Iveland[1], E. Jeffrey[1], Z. Jiang[1], C. Jones[1], D. Kafri[1], K. Kechedzhi[1], T. Khattar[1], S. Kim[1], A. Y. Kitaev[1,4], P. V. Klimov[1], A. R. Klots[1], A. N. Korotkov[1,5], F. Kostritsa[1], J. M. Kreikebaum[1], D. Landhuis[1], P. Laptev[1], K.-M. Lau[1], J. Lee[1], L. Laws[1], W. Liu[1], A. Locharla[1], O. Martin[1], J. R. McClean[1], M. McEwen[1,6], B. Meurer Costa[1], K. C. Miao[1], M. Mohseni[1], S. Montazeri[1], A. Morvan[1], E. Mount[1], W. Mruczkiewicz[1], O. Naaman[1], M. Neeley[1], C. Neill[1], M. Newman[1], T. E. O'Brien[1], A. Opremcak[1], A. Petukhov[1], R. Potter[1], C. Quintana[1], N. C. Rubin[1], N. Saei[1], D. Sank[1], K. Sankaragomathi[1], K. J. Satzinger[1], C. Schuster[1], M. J. Shearn[1], V. Shvarts[1], D. Strain[1], Y. Su[1], M. Szalay[1], G. Vidal[1], B. Villalonga[1], C. Vollgraff-Heidweiller[1], T. White[1], Z. Yao[1], P. Yeh[1], J. Yoo[1], A. Zalcman[1], Y. Zhang[1], N. Zhu[1], H. Neven[1], D. Bacon[1], J. Hilton[1], E. Lucero[1], R. Babbush[1], S. Boixo[1], A. Megrant[1], Y. Chen[1], J. Kelly[1], V. Smelyanskiy[1], D. A. Abanin[1,2]*, P. Roushan[1]*

Inherent symmetry of a quantum system may protect its otherwise fragile states. Leveraging such protection requires testing its robustness against uncontrolled environmental interactions. Using 47 superconducting qubits, we implement the one-dimensional kicked Ising model, which exhibits nonlocal Majorana edge modes (MEMs) with $\mathbb{Z}_2$ parity symmetry. We find that any multiqubit Pauli operator overlapping with the MEMs exhibits a uniform late-time decay rate comparable to single-qubit relaxation rates, irrespective of its size or composition. This characteristic allows us to accurately reconstruct the exponentially localized spatial profiles of the MEMs. Furthermore, the MEMs are found to be resilient against certain symmetry-breaking noise owing to a prethermalization mechanism. Our work elucidates the complex interplay between noise and symmetry-protected edge modes in a solid-state environment.

The symmetry of a quantum system can give rise to topologically distinct degenerate ground states. A quantum superposition of such states is, in principle, immune to dephasing; additionally, an energy gap separates the ground states from the excited states and further protects the ground states from energy decay. As such, symmetry-protected ground states may form decoherence-free subspaces (*1*–*4*) and are promising candidates for topological quantum computing (*5*, *6*). An example model supporting symmetry-protected topological states is the Kitaev model of spinless fermions in a one-dimensional (1D) wire (*7*). The $\mathbb{Z}_2$ parity symmetry of the model leads to a pair of degenerate ground states. The distinct parities of the two ground states protect them against local parity-preserving noise, such as potential fluctuations (*8*). The topological property of these degenerate ground states is commonly described by a pair of localized Majorana edge modes (MEMs) at the ends of the wire.

Whereas the degree of symmetry protection in a closed quantum system is often understood, experimental quantum systems are invariably subject to physical noise sources that do not necessarily respect the underlying symmetry. In the context of MEMs, notable efforts have been directed toward experimentally realizing the Kitaev model, for example, in nanowires with spin-orbit interactions placed in the proximity of a superconductor (*9*–*16*). Here, the underlying $\mathbb{Z}_2$ symmetry cannot be broken by local perturbations within a closed system. Nevertheless, theoretical results have widely suggested that MEMs remain susceptible to a variety of decoherence effects from their open solid-state environment (*17*–*20*). Experimental results have also established that the density of subgap quasiparticles is often orders of magnitude higher than predictions from simple thermal population arguments (*21*–*24*). The incoherent processes involving these quasiparticles can change the parity of the ground state and, consequently, destroy the topological protection. These results highlight the importance of characterizing the extent of symmetry protection in realistic open-system environments.

The advent of high-fidelity quantum processors and simulators suggests an alternative approach to examining the realistic extent of protection for a given symmetry (*25*–*27*). In this study, we use the Jordan-Wigner transformation (JWT) to map the Kitaev model to a transverse Ising spin model (*28*), which is more compatible with a chain of qubits (*29*–*32*). The JWT also maps each MEM, commonly represented by a sum of local Majorana operators in the fermionic chain, to a sum of Pauli spin operators that can be individually characterized on a quantum processor. Given the nonlocal nature of the JWT, the MEMs in the Pauli basis are prone to local symmetry-breaking noise even within a closed system, which distinguishes them from MEMs in fermionic systems. Despite this disadvantage, we find that the interplay between $\mathbb{Z}_2$ parity symmetry and a prethermalization mechanism endows the MEMs with a strong resilience toward both closed-system thermalization and open-system perturbations such as low-frequency noise. Furthermore, we discover a method for accurately reconstructing the Pauli expansion of MEMs in the presence of decoherence, which may be extended to study other integrals of motion in many-body quantum systems.

The experiment is conducted on an open-ended chain of L = 47 superconducting qubits [see (*33*) for device details]. The qubit chain is periodically driven by a quantum circuit corresponding to a kicked Ising model (Fig. 1B),

[1]Google Research, Mountain View, CA, USA. [2]Department of Theoretical Physics, University of Geneva, Geneva, Switzerland. [3]Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA, USA. [4]Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, CA, USA. [5]Department of Electrical and Computer Engineering, University of California, Riverside, CA, USA. [6]Department of Physics, University of California, Santa Barbara, CA, USA.
*Corresponding author. Email: banin@google.com (D.A.A.); pedramr@google.com (P.R.)

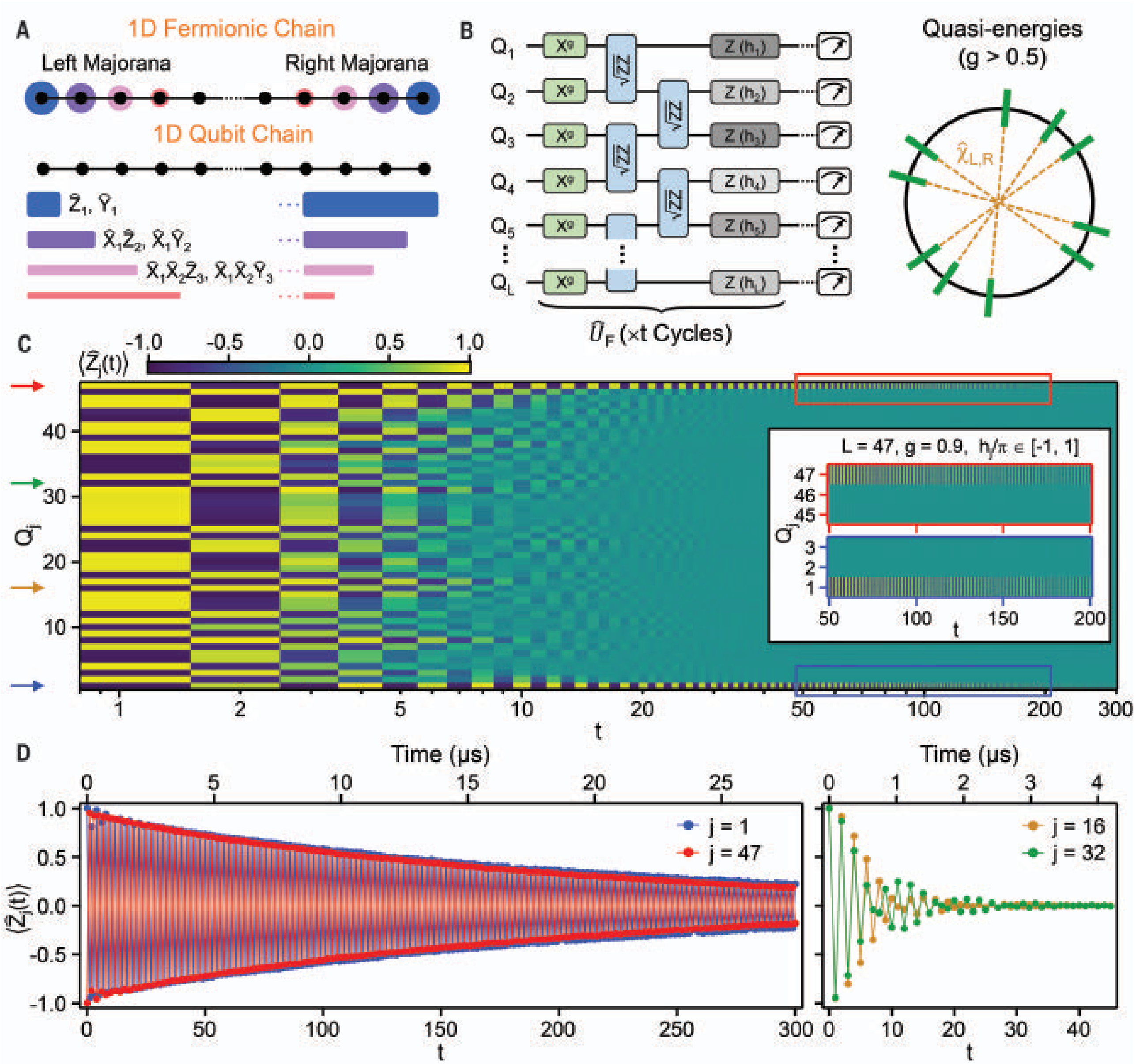


Fig. 1. Observation of long-lived edge modes in a kicked Ising model. (**A**) Schematic illustration of the Jordan-Wigner transformation between a 1D fermionic Kitaev chain and a qubit chain. In the fermionic (qubit) chain, the sizes (widths) of the colored spheres (bars) denote the relative weights of the edge modes in the Majorana fermion (Pauli) basis. The right edge mode in the Pauli basis is dominated by long Pauli operators spanning the entire chain. (**B**) (Left) Quantum circuit implementation of a kicked Ising model. An identical unitary $\hat{U}_F$ is repeated a total of t times. (Right) Eigenstates of $\hat{U}_F$ ($g > 0.5$), shown on a unit circle according to their quasienergies. (**C**) $\langle \hat{Z}_j(t) \rangle$ as a function of t and qubit location Q_j. The initial state is a random product state, $|0101001...\rangle$. (Inset) $\langle \hat{Z}_j(t) \rangle$ for the three leftmost and rightmost qubits, between t = 50 and t = 200. (**D**) $\langle \hat{Z}_j(t) \rangle$ for the two edge qubits j = 1 and 47 (left panel) and two qubits within the bulk j = 16 and 32 (right panel). Top axis for each plot indicates real time, calculated on the basis of the time needed to execute $\hat{U}_F$ (93 ns). Locations for the qubits shown in this panel are also indicated by colored arrows in (C).

with the following unitary applied in each cycle

$$\hat{U}_F = e^{-\frac{i}{2}\sum_{j=1}^{L} h_j \hat{Z}_j} e^{-\frac{i\pi J}{2}\sum_{j=1}^{L-1} \hat{Z}_j \hat{Z}_{j+1}} e^{-\frac{i\pi g}{2}\sum_{j=1}^{L} \hat{X}_j} \quad (1)$$

where $\hat{X}_j$ and $\hat{Z}_j$ denote Pauli operators acting on a given qubit Q_j. Here J and g denote the strengths of the Ising interaction and the transverse fields, respectively; and h_j is a set of local z-fields that break integrability of the model. Compared to a digitized implementation of the transverse Ising Hamiltonian (see fig. S11 for experimental data of this approach), the periodic (i.e., Floquet) evolution here generates faster dynamics in real time and is advantageous given the finite coherence times of the qubits. Under such a drive, the Hilbert space of the system may be described by the eigenstates of $\hat{U}_F$ whose eigenvalues lie on a unit circle, as illustrated in the right panel of Fig. 1B.

Given that there is no ground state in a Floquet system, the twofold degeneracy of the ground state in the Kitaev model instead becomes a pairing of eigenstates across the entire spectrum. In our work, we fix J = 1/2, wherein the Floquet system, in the integrable limit h_j = 0, has $\mathbb{Z}_2$ spin-flip symmetry and exhibits two phases with distinct spectral pairings. For

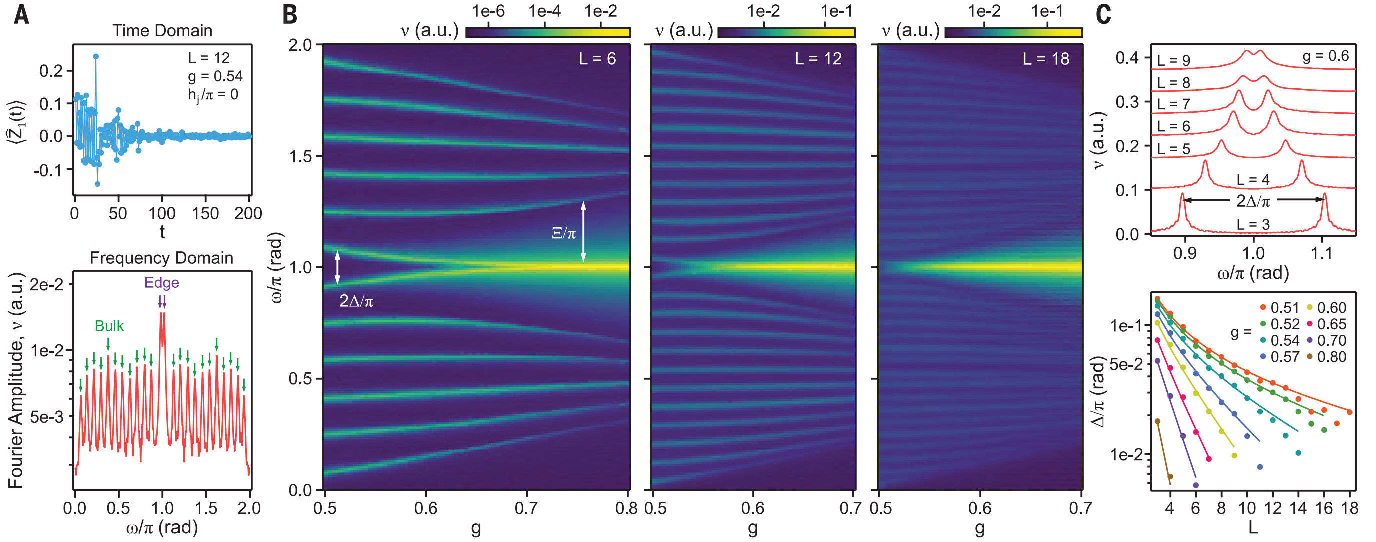


Fig. 2. Quasienergy spectroscopy. (**A**) (Top) $\langle \hat{Z}_1(t) \rangle$ measured in the integrable limit $h_j = 0$ and system size $L = 12$. (Bottom) Frequency-dependent amplitude $\nu(\omega)$ of the Fourier transform of $\langle \hat{Z}_1(t) \rangle$ in the top panel. The arrows indicate the single-particle quasienergy peaks for the bulk and edge fermionic modes. a.u., arbitrary units. (**B**) ν as a function of both frequency ω/π and g, measured for three different values of L. In all cases, $h_j = 0$. To obtain the spectra, $\langle \hat{Z}_1(t) \rangle$ is measured up to t = 300, 200, and 150 cycles for L = 6, 12, and 18, respectively. (**C**) (Top) $\nu(\omega)$ for different L, showing the quasienergy peaks of the hybridized MEMs with a splitting $2\Delta/\pi$. Data are offset for clarity. (Bottom) Δ/π measured as a function of L at different values of g. Solid lines represent exact numerical results from diagonalizing $\hat{U}_F$ in the fermionic basis (*33*). Random product states are used as initial states in all measurements.

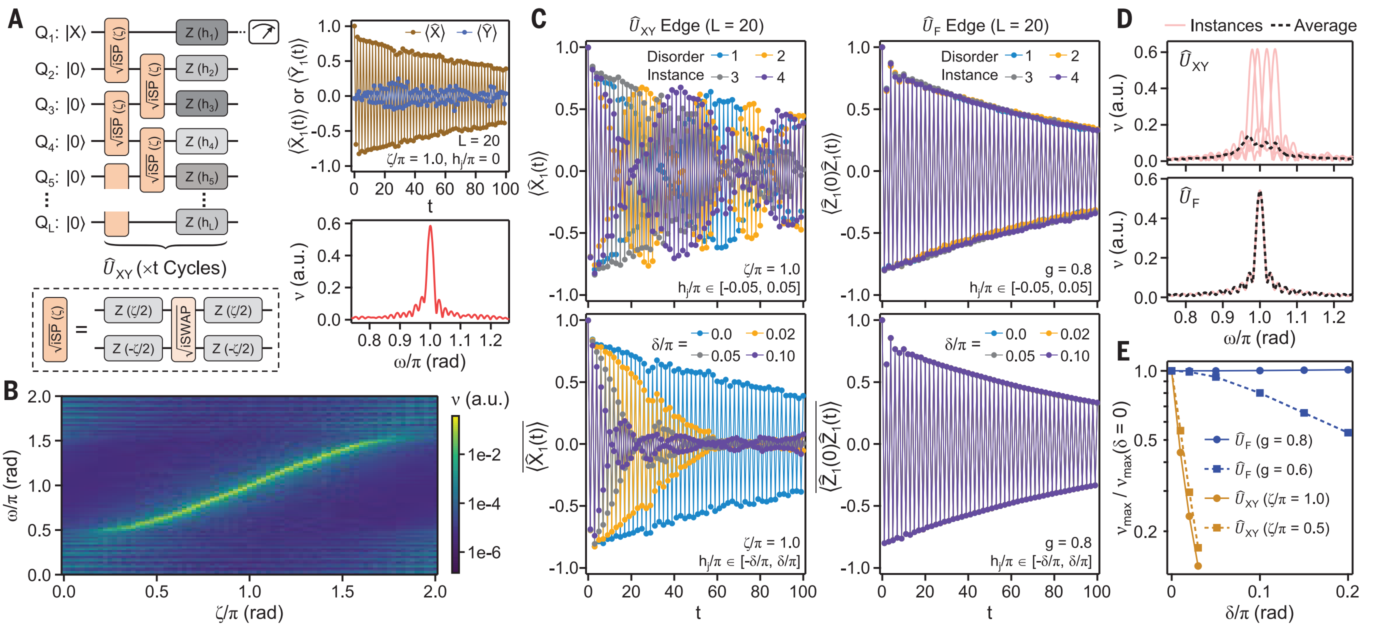


Fig. 3. Low-frequency noise resilience of MEMs and comparison with unprotected edge modes. (**A**) (Left) Quantum circuit corresponding to the XY model, where an identical cycle unitary $\hat{U}_{XY}$ is applied t times. (Right) Top panel shows $\langle \hat{X}_1(t) \rangle$ and $\langle \hat{Y}_1(t) \rangle$ measured at Q_1, with the control parameter $\zeta/\pi = 1.0$ and no disorder $h_j/\pi = 0$. Bottom panel shows the Fourier spectrum $\nu(\omega)$ of $\langle \hat{X}_1(t) \rangle + i\langle \hat{Y}_1(t) \rangle$. (**B**) $\nu(\omega)$ as a function of ω/π and ζ for the $\hat{U}_{XY}$ model, where $\langle \hat{X}_1(t) \rangle$ and $\langle \hat{Y}_1(t) \rangle$ are measured up to t = 100. (**C**) (Top panels) $\langle \hat{X}_1(t) \rangle$ $\left(\langle \hat{Z}_1(0)\hat{Z}_1(t) \rangle\right)$ for the $\hat{U}_{XY}$ ($\hat{U}_F$) edge modes, measured for four different disorder realizations with $h_j/\pi \in [-0.05, 0.05]$. (Bottom panels) Disorder-averaged $\overline{\langle \hat{X}_1(t) \rangle}$ $\left(\overline{\langle \hat{Z}_1(0)\hat{Z}_1(t) \rangle}\right)$ for the $\hat{U}_{XY}$ ($\hat{U}_F$) edge mode, shown over four different disorder strengths δ. Eighty disorder instances $h_j/\pi \in [-\delta/\pi, \delta/\pi]$ are used for averaging in each case, and the initial states are additionally randomized between instances for $\hat{U}_F$. (**D**) Red lines: Fourier spectra $\nu(\omega)$ obtained from the disorder instances in the upper panels of (C). Black lines: $\nu(\omega)$ for the disorder-averaged observables (δ = 0.05) in the lower panels of (C). (**E**) Maximum Fourier amplitude ν_{max} as a function of δ. Data are normalized by ν_{max} at δ = 0.

the phase characterized by $g > 0.5$, which is the focus of the main text, the quasienergy levels have a π-pairing (Fig. 1B, right panel): Every many-body eigenstate of the $\hat{U}_F$ with quasi-energy θ has a "partner" state with quasi-energy θ + π (*34*). The transition between any paired eigenstates is enabled by an application of the so-called π-MEMs, $\hat{\chi}_L$ and $\hat{\chi}_R$ (*35*, *36*). The π-MEMs anticommute with $\hat{U}_F$ in the large L limit

$$\hat{\chi}_L\hat{U}_F = -\hat{U}_F\hat{\chi}_L, \hat{\chi}_R\hat{U}_F = -\hat{U}_F\hat{\chi}_R \qquad (2)$$

At $g < 0.5$, the eigenspectrum of $\hat{U}_F$ has a double degeneracy, that is, each eigenstate has a partner state with the same quasienergy.

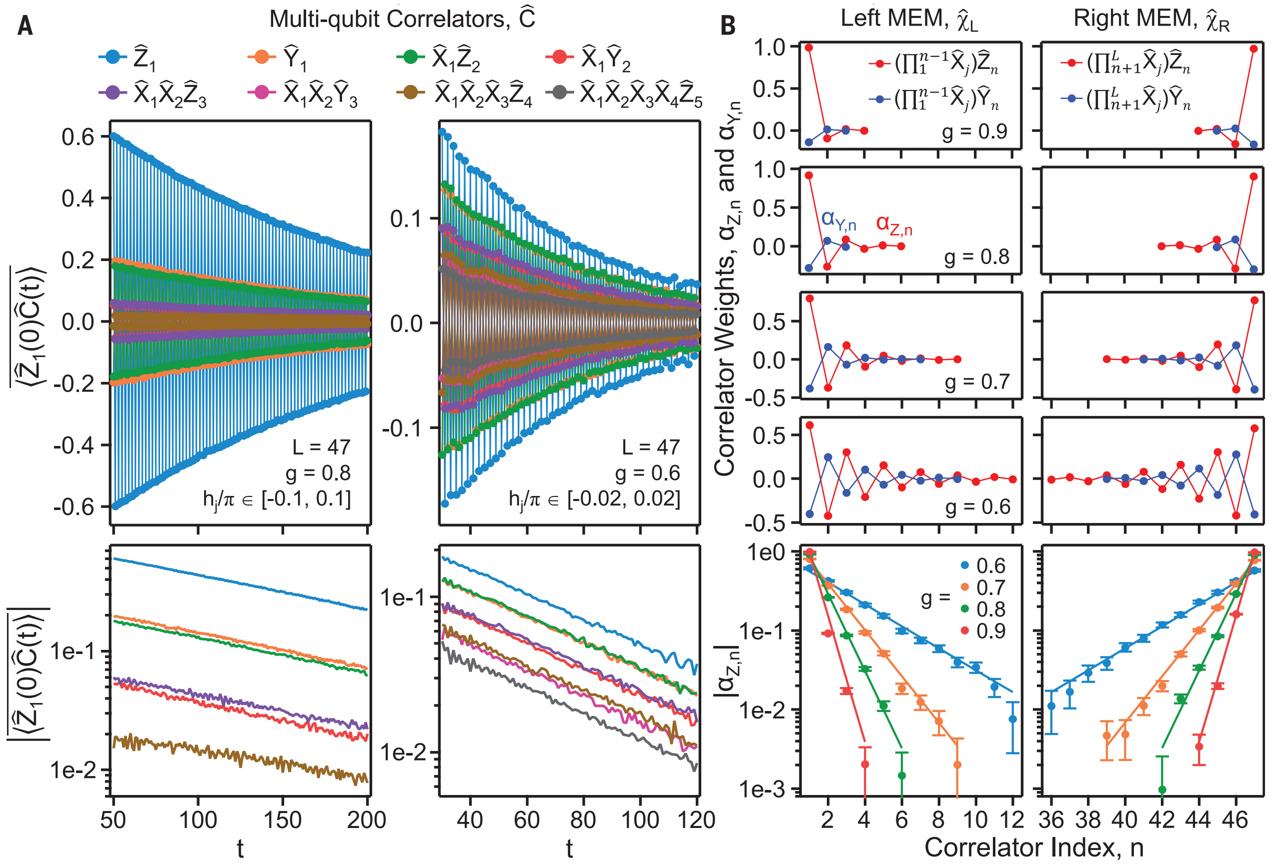


Fig. 4. Reconstructing the Pauli operator expansion of MEMs. (**A**) (Top panels) Correlators $\overline{\langle \hat{Z}_1(0)\hat{C}(t)\rangle}$ for g = 0.8 and g = 0.6, with the compositions of $\hat{C}(t)$ shown in the legend. Here, the g = 0.8 (g = 0.6) data are averaged over 10 (12) disorder realizations and initial random product states. Bottom panels show the absolute values of the correlators, $\left|\overline{\hat{Z}_1(0)\hat{C}(t)}\right|$. (**B**) The top eight panels show experimentally reconstructed Pauli operator expansion of the MEMs $\hat{\chi}_{L,R}$; $\alpha_{Z,n}$ and $\alpha_{Y,n}$ correspond to the coefficients of the Pauli operators shown in the legends. The bottom two panels show experimental values of $|\alpha_{Z,n}|$ (points) and theoretical predictions (solid lines). Error bars correspond to statistical uncertainty stemming from single-shot measurements [see methods (33)].

Here, the transition between paired eigenstates is described by two so-called 0-MEMs that commute with $\hat{U}_F$. Experimental data for this regime, which is analogous to the ferromagnetic phase of the transverse Ising model, are shown in fig. S10. At the critical point $g = 0.5$, the eigenstates are distributed uniformly on the unit circle with a gap of π/L, which vanishes in the limit $L = \infty$.

In the presence of finite local fields $h_j \neq 0$, $\hat{U}_F$ is no longer integrable and the $\mathbb{Z}_2$ symmetry is also broken. We begin by searching for signatures of stable edge modes in this regime, focusing on the $\hat{Z}$ operators that, in the JWT, have large overlap with MEMs on the edge (Fig. 1A). Figure 1C shows experimental measurements of $\langle \hat{Z}_j(t)\rangle$ for all qubits in the chain, where we have chosen h_j/π from a random uniform distribution [–1,1] to maximize the effect of integrability breaking. We observe a stark contrast in the behavior of the edge qubits, Q_1 and Q_{47}, and qubits within the chain, Q_2 to Q_{46}. Whereas $\langle \hat{Z}_j(t)\rangle$ decays rapidly to 0 after ~20 cycles (~2 µs) for any qubit in the bulk, $\langle \hat{Z}_j(t)\rangle$ decays much more slowly for the edge qubits. In addition, $\langle \hat{Z}_j(t)\rangle$ for each edge qubit shows a subharmonic oscillation at a period twice that of the drive $\hat{U}_F$, because each application of $\hat{U}_F$ changes the sign of $\hat{\chi}_{L,R}$ owing to their anticommutation (Eq. 2).

The bulk-edge difference is further illustrated in Fig. 1D, where data for four qubits are shown. The lifetimes of the edge modes, which include contributions from both external decoherence effects and internal nonintegrable dynamics, are extracted by fitting the envelope of $\langle \hat{Z}_1(t)\rangle$ ($\langle \hat{Z}_{47}(t)\rangle$) to an exponential (fig. S8) and found to be 19.5 µs (17.2 µs) for Q_1 (Q_{47}). These values are close to the typical single-qubit relaxation time T_1 = 22.2 µs on the device—a preliminary indication that the MEMs are resilient toward integrability- and symmetry-breaking fields as well as dephasing effects such as low-frequency noise.

Recent theoretical works have suggested that the resilience of the edge modes toward nonintegrable dynamics is a result of prethermalization (37–41). Unlike thermalizing systems, which monotonically decay to ergodic states over time, a prethermal system relaxes first to a metastable state before decaying to ergodic states. A common mechanism for prethermalization is the existence of spectral gaps, which make relaxation processes driven by integrability-breaking perturbations off-resonant, thereby preventing energy absorption. To experimentally establish prethermalization in our system, we characterize the excitation spectrum in the integrable limit, $h_j = 0$. Here, the many-body spectrum of $\hat{U}_F$ may be constructed from a total of $2L$ energy quanta, corresponding to the quasienergies of noninteracting Bogoliubov fermionic quasiparticles in the fermionic representation of $\hat{U}_F$. These quasienergies can be obtained through a Fourier analysis of time-domain signals (42, 43) [see supplementary text sections III and V (33)]. Figure 2A shows measurements of $\langle \hat{Z}_1(t)\rangle (h_j = 0)$ for a short chain L = 12. The time evolution for $\langle \hat{Z}_1(t)\rangle$ is now seemingly featureless, which results from interference between different eigenmodes of $\hat{U}_F$. To obtain the quasienergies, a Fourier transform of the time-domain data is then performed (Fig. 2A, bottom). The Fourier spectrum $\nu(\omega)$ reveals a total of $2L$ distinct peaks at values of

ω corresponding to the quasienergies of the $2L$ noninteracting fermionic modes in the system.

The two dominant peaks close to $\omega = \pi$ in the spectrum of Fig. 2A are associated with the MEMs, which are split in quasienergy because of their hybridization in this short chain. To confirm this interpretation, we change the localization length ξ of the MEMs by tuning g and measure the spectra over three different system sizes. The results, shown in Fig. 2B, reveal two important features: First, we observe that the quasienergy splitting 2Δ of the two MEMs decreases as g increases. This is caused by a reduced ξ that leads to weaker hybridization between $\hat{\chi}_L$ and $\hat{\chi}_R$. Second, we observe a finite quasienergy gap Ξ between the MEMs and the other bulk fermionic modes, which increases at larger g. This quasienergy gap, which crucially remains open as L increases, suppresses transitions between bulk and edge states and is the key to protecting the MEMs against integrability-breaking fields. Further discussion of this prethermalization mechanism is presented in supplementary text section IV (*33*).

Although the bulk gap protects the MEMs from internal thermalization, the finite quasienergy difference 2Δ between the two MEMs is sensitive to disorder fluctuations. Such a sensitivity may lead to dephasing of the MEMs through low-frequency noise (*20*), as shown by fig. S5. This effect may be suppressed by reducing the hybridization between $\hat{\chi}_L$ and $\hat{\chi}_R$, which is achievable through increasing either g (Fig. 2B) or L. The dependence of Δ on L is mapped out in detail by the experimental measurements shown in Fig. 2C. We observe that for $g > 0.6$, Δ is exponentially suppressed by larger L, in agreement with theory (*33*). For $g < 0.6$, the suppression is no longer exponential, given the proximity to the phase transition point $g = 0.5$, where the bulk gap closes. We also find excellent agreement between exact numerical results and experimental measurements even for $\Delta/\pi \approx 0.01$, which is a result of accurate gate calibrations described in (*33*).

We next perform a systematic study on the low-frequency noise resilience of the MEMs for a moderately long chain, $L = 20$. To examine the role of symmetry in noise protection, we have also experimentally realized edge modes in a different periodic circuit with a cycle unitary $\hat{U}_{XY}$ that does not have $\mathbb{Z}_2$ symmetry. As illustrated in Fig. 3A, $\hat{U}_{XY}$ consists of two layers of two-qubit gates applied between all nearest-neighbor qubits, $\sqrt{\mathrm{iSP}(\zeta)} = e^{-i\frac{\zeta}{4}\left(\hat{Z}_j-\hat{Z}_{j+1}\right)} e^{-i\frac{\pi}{4}\left(\hat{\sigma}_j^+\hat{\sigma}_{j+1}^- + \hat{\sigma}_j^-\hat{\sigma}_{j+1}^+\right)} e^{-i\frac{\zeta}{4}\left(\hat{Z}_j-\hat{Z}_{j+1}\right)}$, where $\hat{\sigma}^{+,-}$ denotes Pauli raising and lowering operators. In the single-excitation subspace, $\hat{U}_{XY}$ has L eigenmodes, including two localized edge modes [see appendix E of (*44*)] for control parameter $\zeta/\pi \in [0.25, 1.75]$. The leading order terms in the Pauli operator expansion of the edge modes are $\hat{\sigma}_1^+$ and $\hat{\sigma}_L^+$, respectively.

To probe one of the edge modes for $\hat{U}_{XY}$, we prepare the system in a superposition state $\frac{1}{\sqrt{2}}(|0000\ldots\rangle + |1000\ldots\rangle)$ to maximize the initial value $\langle\hat{\sigma}_1^+(t=0)\rangle = 1$. We then apply $\hat{U}_{XY}$ t times before measuring the time-dependent observable $\langle\hat{\sigma}_1^+(t)\rangle = \langle\hat{X}_1(t)\rangle + i\langle\hat{Y}_1(t)\rangle$, which precesses at a frequency corresponding to the quasienergy of the edge mode. Example experiment data and the corresponding Fourier spectrum $\nu(\omega)$ for $\zeta/\pi = 1.0$ are both shown in Fig. 3A, demonstrating a slowly decaying subharmonic response and a quasienergy peak at $\omega = \pi$ that are similar to those of the MEMs of the $\hat{U}_F$ model. Figure 3B shows experimentally measured ν as a function of both ζ and ω. At $0.25 \lesssim \zeta/\pi \lesssim 1.75$, we observe a dominant quasienergy peak that corresponds to an edge mode and is separated from the $L - 2$ bulk modes, visible as smaller peaks outside the range $0.5 < \omega_\pi < 1.5$, by spectral gaps akin to the bulk gap Ξ of the $\hat{U}_F$ model. Despite these apparent similarities, a crucial distinction exists between the two models: The quasienergy of the $\hat{U}_{XY}$ edge mode is first-order sensitive to ζ at all values of ζ, whereas the quasienergy of the $\hat{U}_F$ edge mode asymptotically approaches π as g increases. This distinction stems from the lack of $\mathbb{Z}_2$ symmetry in $\hat{U}_{XY}$ and leads to drastically different robustness of the two models toward low-frequency noise in h_j, which we explore next.

The upper panels of Fig. 3C show $\langle\hat{X}_1(t)\rangle$ $\left(\langle\hat{Z}_1(0)\hat{Z}_1(t)\rangle\right)$ of the $\hat{U}_{XY}\left(\hat{U}_F\right)$ model, measured for four different realizations of h_j/π that are uniformly chosen from $[-\delta, \delta]$. Here, the autocorrelator $\langle\hat{Z}_1(0)\hat{Z}_1(t)\rangle$ differs from $\langle\hat{Z}_1(t)\rangle$ only by a random ± sign given by the initial state of Q_1, and $\delta = 0.05$ is a disorder strength chosen to be comparable to the low-frequency fluctuation of the quantum device. We observe that $\langle\hat{X}_1(t)\rangle$ exhibits beating patterns that depend sensitively on the disorder realization. This is a result of the first-order sensitivity toward control parameters demonstrated in Fig. 3B. On the other hand, $\langle\hat{Z}_1(0)\hat{Z}_1(t)\rangle$ is virtually unchanged between different disorder realizations. The impact of low-frequency noise on each edge mode realization is then emulated by averaging the corresponding observable over an ensemble of disorder realizations, which mimics the process of dephasing. The disorder-averaged $\overline{\langle\hat{X}_1(t)\rangle}$ in the $\hat{U}_{XY}$ model, shown in the lower panel of Fig. 3C, decays significantly faster as the disorder strength δ increases. On the other hand, $\overline{\langle\hat{Z}_1(0)\hat{Z}_1(t)\rangle}$ in the $\hat{U}_F$ model remains unchanged over δ.

The sensitivity of the two edge mode realizations toward low-frequency noise is further elucidated by inspecting the Fourier spectrum $\nu(\omega)$ of each disorder realization, shown in Fig. 3D. Here, we observe that the quasienergy peak for the $\hat{U}_{XY}$ edge mode is different for each disorder realization, resulting in a broadened spectrum with a lower peak height upon averaging. On the other hand, the quasienergy peak for $\hat{U}_F$ remains stable at $\omega = \pi$, irrespective of disorder realizations. Lastly, we measure the disorder-averaged quasienergy peak height, $\nu_{max} = \mathrm{Max}[\nu(\omega)]$, and show the results in Fig. 3E. For the $\hat{U}_{XY}$ model, we observe that ν_{max} decays exponentially as a function of δ irrespective of ζ. For the $\hat{U}_F$ model, ν_{max} is completely insensitive to δ for sufficiently localized MEMs ($g = 0.8$) and remains insensitive for small $\delta < 0.05$ even in the more delocalized regime $g = 0.6$. These results highlight the critical role of symmetry in stabilizing the quasienergies of MEMs and protecting their lifetimes against low-frequency noise.

Finally, using the full $L = 47$ qubit chain, we demonstrate an error-mitigation strategy for accurately reconstructing the Pauli operator expansion of $\hat{\chi}_{L,R}$ in the presence of noise. Figure 4A shows the late-time evolution of eight multiqubit Pauli operators $\hat{C}$ entering the JWT of $\hat{\chi}_{L,R}$, experimentally obtained by rotating each qubit 1into the appropriate basis followed by multiqubit readout. We observe that each $\overline{\hat{Z}_1(0)\hat{C}(t)}$ exhibits a similar subharmonic response, with an amplitude that decreases when $\hat{C}$ incorporates more qubits and has less overlap with $\hat{\chi}_{L,R}$ (Fig. 1A). Operators ending with $\hat{Y}$ also show smaller amplitudes than those ending with $\hat{Z}$ because they have no overlap with $\hat{\chi}_{L,R}$ in the time-independent transverse Ising model and only arise as corrections to the JWT of $\hat{\chi}_{L,R}$ as a consequence of the time-dependent, periodic dynamics. Notably, as shown also in Fig. 4A, the absolute values (i.e., magnitudes) of these operators, $\left|\overline{\hat{Z}_1(0)\hat{C}(t)}\right|$, exhibit nearly identical decay rates despite their different lengths and compositions.

The observation in Fig. 4A is contrary to naïve expectations, wherein the decay rate of a quantum operator is expected to scale with the number of qubits it incorporates. The result may be qualitatively understood by the fact that $\hat{\chi}_L$ and $\hat{\chi}_R$ anticommute with $\hat{U}_F$ (Eq. 2) and are conserved under the periodic dynamics. Even though external decoherence and integrability-breaking fields violate this commutation, $\hat{\chi}_{L,R}$ remains a slowly decaying mode [see supplementary text section VII (*33*)]. As a result, any multiqubit operator having a finite overlap with $\hat{\chi}_{L,R}$ will exhibit a slow-decaying expectation value in its late-time dynamics, with an amplitude proportional to the overlap.

The uniform decay rates of $\overline{\hat{Z}_1(0)\hat{C}(t)}$ inform an experimental strategy for reconstructing the expansion of the MEMs in the Pauli operator basis

$$\hat{\chi}_{L,R} = \sum_{n=1}^{L}\left[\alpha_{Z,n}\left(\prod_{j=M}^{N}\hat{X}_j\right)\hat{Z}_n + \alpha_{Y,n}\left(\prod_{j=M}^{N}\hat{X}_j\right)\hat{Y}_n\right] \quad (3)$$

where the products over j have limits $N = n - 1$ and $M = 1$ for $\hat{\chi}_L$, and $N = L$ and $M = n + 1$ for $\hat{\chi}_R$

(*45*). The coefficients $\alpha_{Z,n}$ and $\alpha_{Y,n}$ normalize to unity for $L = \infty$: $\sum_n |\alpha_{Z,n}|^2 + |\alpha_{Y,n}|^2 = 1$. To estimate their values, we measure different $\hat{Z}_1(0)\hat{C}(t)$ at 10 late-time cycles. The average value of each operator and the normalization condition allow us to determine the ideal values of $\alpha_{Z,n}$ and $\alpha_{Y,n}$ [see methods and supplementary text section VI (*33*) for details].

The experimentally measured coefficients, shown in Fig. 4B for four values of g, both oscillate in sign and decay exponentially as n moves away from the edge. The decay rate is also observed to decrease as g approaches the critical value $g = 0.5$. This is a result of the fact that the decay constant for the coefficients is the localization length ξ of the MEMs, which diverges at $g = 0.5$ [see supplementary text section III (*33*)]. A comparison between theoretical and experimental values of $|\alpha_{Z,n}|$ is shown in the bottom panels of Fig. 4B, where good agreement is found over a span of nearly three orders of magnitude.

In this study, we simulate MEMs using a system of driven transmon qubits and comprehensively study their symmetry protection against noise in their solid-state environment. We find that the degree of protection sensitively depends on the physical characteristic of the noise and generally does not extend to noise that breaks the underlying symmetry, such as T_1 decay of the transmon qubits. We also find that, because of a prethermalization mechanism, the MEMs in our system are protected against certain noise that seemingly violates $\mathbb{Z}_2$ symmetry, for example, local $\hat{Z}$ fluctuations. These results highlight the complex interplay between physical noise and protection and indicate the crucial importance of testing symmetry against open-system dynamics in any experimental platform. Furthermore, we find that even in the presence of decoherence, the Pauli expansion of conserved quantities such as MEMs can be accurately determined by measuring and renormalizing late-time expectation values of Pauli operators. This error-mitigation strategy may be applied to study integrals of motion in physical models that are more difficult to compute classically. Preliminary results on nonintegrable dynamics are shown in fig. S7.

REFERENCES AND NOTES

1. P. Zanardi, M. Rasetti, *Phys. Rev. Lett.* **79**, 3306–3309 (1997).
2. D. A. Lidar, I. L. Chuang, K. B. Whaley, *Phys. Rev. Lett.* **81**, 2594–2597 (1998).
3. D. Bacon, J. Kempe, D. A. Lidar, K. B. Whaley, *Phys. Rev. Lett.* **85**, 1758–1761 (2000).
4. A. Y. Kitaev, *Ann. Phys.* **303**, 2–30 (2003).
5. C. Nayak, S. H. Simon, A. Stern, M. Freedman, S. Das Sarma, *Rev. Mod. Phys.* **80**, 1083–1159 (2008).
6. A. G. Fowler, M. Mariantoni, J. M. Martinis, A. N. Cleland, *Phys. Rev. A* **86**, 032324 (2012).
7. A. Y. Kitaev, *Phys. Uspekhi* **44**, 131–136 (2001).
8. N. Read, D. Green, *Phys. Rev. B Condens. Matter* **61**, 10267–10297 (2000).
9. R. M. Lutchyn, J. D. Sau, S. Das Sarma, *Phys. Rev. Lett.* **105**, 077001 (2010).
10. Y. Oreg, G. Refael, F. von Oppen, *Phys. Rev. Lett.* **105**, 177002 (2010).
11. V. Mourik *et al.*, *Science* **336**, 1003–1007 (2012).
12. L. P. Rokhinson, X. Liu, J. K. Furdyna, . *Nat. Phys.* **8**, 795–799 (2012).
13. A. Das *et al.*, *Nat. Phys.* **8**, 887–895 (2012).
14. S. Nadj-Perge *et al.*, *Science* **346**, 602–607 (2014).
15. S. M. Albrecht *et al.*, *Nature* **531**, 206–209 (2016).
16. P. Yu *et al.*, *Nat. Phys.* **17**, 482–488 (2021).
17. G. Goldstein, C. Chamon, *Phys. Rev. B Condens. Matter Mater. Phys.* **84**, 205109 (2011).
18. M. Cheng, R. M. Lutchyn, S. Das Sarma, *Phys. Rev. B Condens. Matter Mater. Phys.* **85**, 165124 (2012).
19. J. C. Budich, S. Walter, B. Trauzettel, *Phys. Rev. B Condens. Matter Mater. Phys.* **85**, 121405 (2012).
20. C. Knapp, T. Karzig, R. M. Lutchyn, C. Nayak, *Phys. Rev. B* **97**, 125404 (2018).
21. E. M. Levenson-Falk, F. Kos, R. Vijay, L. Glazman, I. Siddiqi, *Phys. Rev. Lett.* **112**, 047002 (2014).
22. B. Feldman *et al.*, *Nat. Phys.* **13**, 286–291 (2016).
23. K. Serniak *et al.*, *Phys. Rev. Lett.* **121**, 157701 (2018).
24. M. Hays *et al.*, *Phys. Rev. Lett.* **121**, 047001 (2018).
25. R. Blatt, C. F. Roos, *Nat. Phys.* **8**, 277–284 (2012).
26. C. Gross, I. Bloch, *Science* **357**, 995–1001 (2017).
27. I. Carusotto *et al.*, *Nat. Phys.* **16**, 268–279 (2020).
28. E. Lieb, T. Schultz, D. Mattis, *Ann. Phys.* **16**, 407–466 (1961).
29. L. S. Levitov, T. P. Orlando, J. B. Majer, J. E. Mooij, arXiv:cond-mat/0108266 [cond-mat.mes-hall] (2001).
30. J. Q. You, Z. D. Wang, W. Zhang, F. Nori, *Sci. Rep.* **4**, 5535 (2014).
31. J.-S. Xu *et al.*, *Sci. Adv.* **4**, eaat6533 (2018).
32. M. I. Dykman, *Phys. Rev. A* **100**, 042101 (2019).
33. Materials and methods are available as supplementary materials.
34. V. Khemani, A. Lazarides, R. Moessner, S. L. Sondhi, *Phys. Rev. Lett.* **116**, 250401 (2016).
35. M. Thakurathi, A. A. Patel, D. Sen, A. Dutta, *Phys. Rev. B Condens. Matter Mater. Phys.* **88**, 155133 (2013).
36. D. J. Yates, F. H. L. Essler, A. Mitra, *Phys. Rev. B* **99**, 205419 (2019).
37. P. Fendley, *J. Phys. A Math. Theor.* **49**, 30LT01 (2016).
38. D. Abanin, W. De Roeck, W. W. Ho, F. Huveneers, *Commun. Math. Phys.* **354**, 809–827 (2017).
39. T. Mori, T. Kuwahara, K. Saito, *Phys. Rev. Lett.* **116**, 120401 (2016).
40. D. V. Else, B. Bauer, C. Nayak, *Phys. Rev. X* **7**, 011026 (2017).
41. D. V. Else, P. Fendley, J. Kemp, C. Nayak, *Phys. Rev. X* **7**, 041062 (2017).
42. P. Roushan *et al.*, *Science* **358**, 1175–1179 (2017).
43. N. Harle, O. Shtanko, R. Movassagh, arXiv:2203.15083 [quant-ph] (2022).
44. C. Neill *et al.*, *Nature* **594**, 508–512 (2021).
45. In Eq. 3, product terms $\prod_{j=M}^{N} \hat{X}_j$ with $N < M$ should be treated as identity $\hat{I}$. Additionally, the JWT of $\hat{\chi}_R$ should, in principle, have higher weights in longer Pauli strings, with the dominant string being $\left(\prod_{j=1}^{L-1} \hat{X}_j\right)\hat{Y}_L$. This is equivalent to $\hat{Z}_L$ up to the parity transformation $\prod_{j=1}^{L} \hat{X}_j$, which commutes with $\hat{U}_F$. In Eq. 3, $\hat{\chi}_R$ is expressed in this alternative form and has a leading term $\alpha_{Z,L}\hat{Z}_L$. Such a basis requires measurements of shorter Pauli strings.
46. Google Quantum AI, Data and Simulation Code for "Noise Resilient Edge Modes on a Chain of Superconducting Qubits," version 1, Zenodo (2022); https://doi.org/10.5281/zenodo.7190882.

ACKNOWLEDGMENTS

We have benefited from discussions with M. H. Devoret, L. G. Dias, I. K. Drozdov, P. Ghaemi, and A. Rahmani. **Funding:** D.B. is a CIFAR Associate Fellow in the Quantum Information Science Program. **Author contributions:** D.A.A. and V.Sm. conceived of the project. X.M., D.A.A., and V.Sm. designed the experiment. X.M. executed the experiment. P.R., X.M., D.A.A., M.So., and M.Y.N. performed analysis of the experimental results. K.W.L. and B.F. contributed to measurements in the supplementary materials. X.M., P.R., and D.A.A. wrote the manuscript. V.Sm. and P.R. led and coordinated the project. Infrastructure support was provided by Google Quantum AI. All authors contributed to revising the manuscript and the supplementary materials. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Experimental data shown in the main text and supplementary materials, as well as simulation software code, are available in Zenodo (*46*).

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq5769
Materials and Methods
Supplementary Text
Figs. S1 to S15
References (*47–60*)

Submitted 17 April 2022; accepted 21 October 2022
10.1126/science.abq5769

CHEMICAL PHYSICS

Cavity-enabled enhancement of ultrafast intramolecular vibrational redistribution over pseudorotation

Teng-Teng Chen[1]†, Matthew Du[1]†, Zimo Yang[2]†, Joel Yuen-Zhou[1]*, Wei Xiong[1,2,3]*

Vibrational strong coupling (VSC) between molecular vibrations and microcavity photons yields a few polaritons (light-matter modes) and many dark modes (with negligible photonic character). Although VSC is reported to alter thermally activated chemical reactions, its mechanisms remain opaque. To elucidate this problem, we followed ultrafast dynamics of a simple unimolecular vibrational energy exchange in iron pentacarbonyl [$Fe(CO)_5$] under VSC, which showed two competing channels: pseudorotation and intramolecular vibrational-energy redistribution (IVR). We found that under polariton excitation, energy exchange was overall accelerated, with IVR becoming faster and pseudorotation being slowed down. However, dark-mode excitation revealed unchanged dynamics compared with those outside of the cavity, with pseudorotation dominating. Thus, despite controversies around thermally activated VSC modified chemistry, our work shows that VSC can indeed alter chemistry through a nonequilibrium preparation of polaritons.

Vibrational strong coupling (VSC) gives rise to delocalized superpositions of molecular vibrations and electromagnetic modes (cavity modes), known as molecular vibrational polaritons (*1*, *2*). Recently, VSC has become a promising method for manipulating chemical reactions in condensed phases (*3–5*). Extensive experimental evidence has shown that without photoexcitation, reaction rates can be either accelerated or decelerated by VSC and reaction selectivity can even be altered (*3–5*). Although much effort

has been devoted to providing a sound explanation for VSC-modified chemistry, consensus between theory and experiments is still missing (*6*–*11*). Though it is clear that polaritons are different from bare molecular states, and thereby have the potential to modify chemistry, dark modes, which greatly outnumber polaritons, have the same excitation energies as the uncoupled vibrational modes (*12*). Hence, some theoretical work predicts that reactivity in the collective VSC regime is similar to that outside an optical cavity (*6*, *8*, *9*, *13*), which is consistent with a few recent experimental results that show no modification of reactions from VSC (*10*, *14*). Resolution of the discrepancies is hindered by the following factors: Most reactions studied so far are quite complex, that is, involve multiple steps or are diffusion limited; and reactive events involving dark modes and polaritons simultaneously are probed together (*3*–*5*). To delineate the effect of VSC, it is therefore critical to study elementary reactions and use a technique that can differentiate the contributions from polaritons and dark modes.

Here, we used ultrafast two-dimensional infrared (2D IR) spectroscopy to follow how the polaritons and dark modes evolve in both pseudorotation and intramolecular vibrational-energy redistribution (IVR) of $Fe(CO)_5$ (*15*). We did so in a state-resolved manner, thereby meeting the desired criteria listed earlier. $Fe(CO)_5$ features two IR-active vibrational bands, a doubly degenerate e′ mode at 1999 cm^{-1} that involves three equatorial CO groups, and an a_2'' mode at 2022 cm^{-1} that involves the axial CO groups (fig. S1). Harris and co-workers applied 2D IR spectroscopy and showed that $Fe(CO)_5$ can rearrange from its D_{3h} equilibrium geometry to a C_{4v} transition state and back to D_{3h}, during which the equatorial and axial CO ligands interconvert, leading to vibrational energy exchange between the a_2'' and e′ modes (*16*). This process, referred to as Berry's pseudorotation, is a single barrier crossing and thus represents the essence of elementary reactions, although the product is indistinguishable from the reactant (*17*). Given that Berry's pseudorotation competes with IVR between a_2'' and e′ modes whose transition dipoles are perpendicular to each other (Fig. 1A, top), $Fe(CO)_5$ is an ideal testbed to understand how VSC affects single barrier-crossing events and the branching ratio between various dynamical processes.

Using 2D IR spectroscopy, we found that when polaritons were pumped, they could accelerate the overall 2D IR cross-peak dynamics, and more notably, the excited polaritons make IVR faster than pseudorotation (bottom of Fig. 1A). By contrast, the dynamics triggered by exciting the dark reservoir modes were similar to the dynamics of molecules outside the cavity. Thus, the fundamental concept of VSC-modified chemistry—that polaritons can change reactions—holds. However, because dark modes are statistically dominant, the overall influence

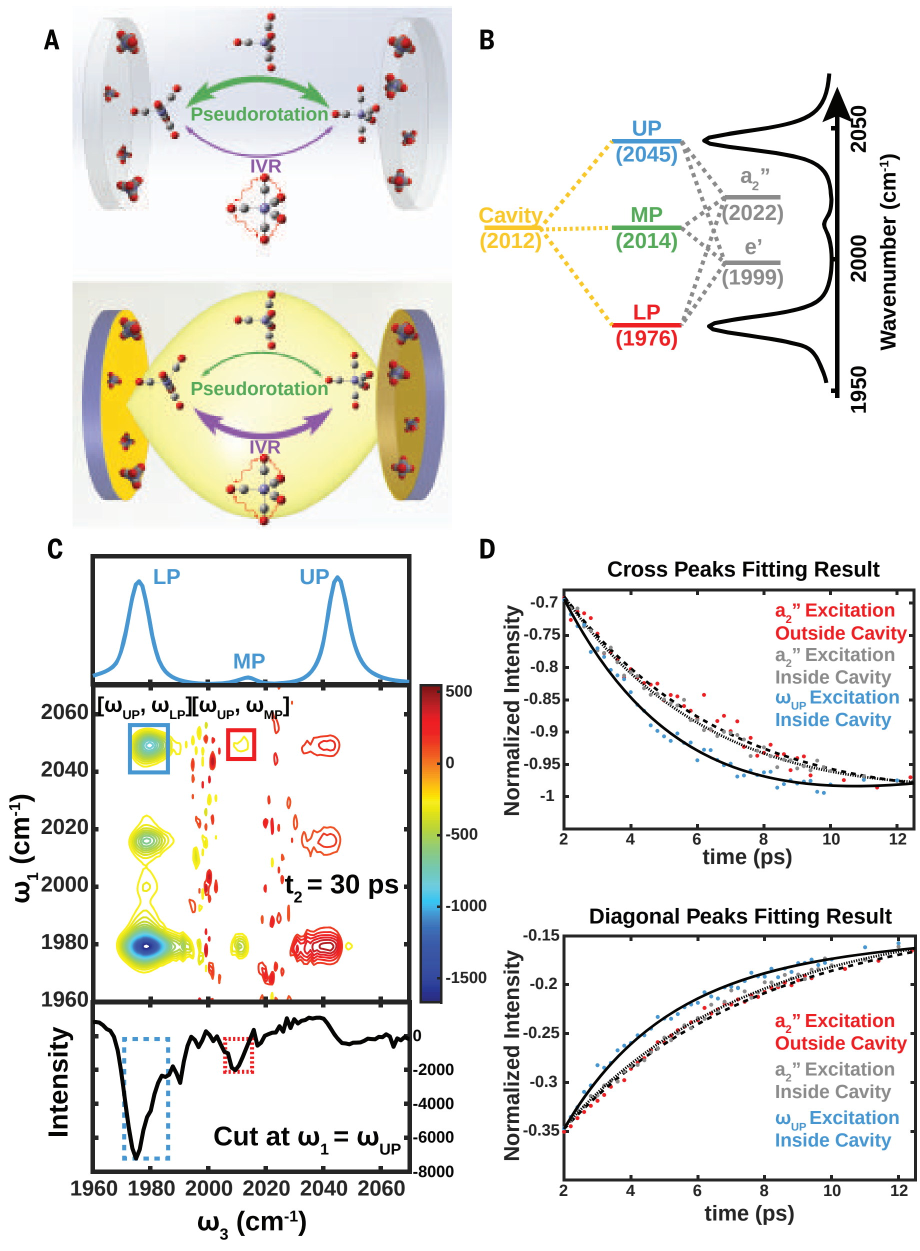


Fig. 1. Influence of VSC on $Fe(CO)_5$ energy-exchange dynamics. (**A**) Schematic drawing showing that when $Fe(CO)_5$ is outside of the cavity, pseudorotation is the dominating channel (top); when the molecule is placed in an optical cavity, IVR becomes the dominant energy-exchange process and pseudorotation is suppressed (bottom). (**B**) Strong coupling diagram and IR spectrum of $Fe(CO)_5$ inside the cavity. (**C**) Normalized 2D IR spectrum using the linear spectrum of strongly coupled $Fe(CO)_5$ at waiting time (t_2) = 30 ps in dodecane (blue and red boxes represent [ω_{UP}, ω_{LP}] and [ω_{UP}, ω_{MP}] peaks, respectively), along with the corresponding linear spectrum (top) and normalized narrowband pump probe spectrum at $\omega_1 = \omega_{UP}$ (bottom). (**D**) Experimental dynamics of cross-peaks (top) and diagonal peaks (bottom) for $Fe(CO)_5$ outside the cavity upon pumping of the a_2'' modes (red dots) and inside the cavity upon pumping of the UP (blue dots) and the a_2'' dark modes (gray dots). The black dashed, dotted, and solid lines are the corresponding fits. Energy is exchanged at a faster rate when pumping the UP, whereas pumping the a_2'' dark modes leads to a rate similar to that outside the cavity.

[1]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA, USA. [2]Materials Science and Engineering Program, University of California, San Diego, La Jolla, CA, USA. [3]Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA, USA.
*Corresponding author. Email: joelyuen@ucsd.edu (J.Y.-Z.); w2xiong@ucsd.edu (W.X.)
†These authors contributed equally to this work.

of VSC on the dynamics of $Fe(CO)_5$, when measured without differentiating polariton and dark modes (e.g., without optical pumping at room temperature), should be negligible.

The VSC condition was achieved by placing a solution of $Fe(CO)_5$ in *n*-dodecane into a Fabry-Pérot microcavity. Unless specifically mentioned, we set the $Fe(CO)_5$ concentration to ~180 mM and the cavity longitudinal thickness to ~12.5 μm. The e′ and a_2'' vibrational modes of $Fe(CO)_5$ strongly couple to a fifth-order cavity mode. The IR spectrum (Fig. 1B) shows the transitions of upper, middle, and lower polaritons (UP, MP, and LP) at a resonant frequency (ω) of ω_{UP} = 2045 cm^{-1}, ω_{MP} = 2014 cm^{-1}, and ω_{LP} = 1976 cm^{-1}, respectively. By fitting to a coupled oscillator model [see supplementary materials (SM) section 2.1], we determined that the cavity mode is 2013 cm^{-1} and it interacts with the e′ and a_2'' modes with amplitudes $g_{cav\text{-}e'}$ = 26 cm^{-1} and $g_{cav\text{-}a2''}$ = 19 cm^{-1}, respectively. Because the full widths at half maximum of the e′ and a_2'' modes are 8 and 5 cm^{-1}, respectively, and that of the cavity mode is 11 cm^{-1}, the samples satisfy the criteria for VSC (*18*).

We used 2D IR to monitor the cross-peak dynamics both outside and inside the cavity (*19*–*22*). Through pseudorotation and IVR, the a_2'' and e′ modes can exchange energy. By measuring the dynamics of the [2022, 1986] cross and [2022, 2010] diagonal peaks of the 1→2 transitions (fig. S6) and fitting them to a kinetic model (see SM section 2.3.1), we determined the energy-exchange rate constant k_{ex} to be 0.084 ± 0.002 ps^{-1} at 25°C (Fig. 1D). Here, unless specifically mentioned, all measurements were done under magic angle conditions to remove contributions from rotational dynamics.

Similarly, for $Fe(CO)_5$ under VSC (Fig. 1C), we followed the dynamics of the [ω_{UP}, ω_{MP}] (red box in Fig. 1C) and [ω_{UP}, ω_{LP}] (blue box in Fig. 1C) peaks of the 2D IR spectrum (or the corresponding narrowband pump probe spectra; see SM section 1.3 for details). Here, we specifically focused on the dynamics involving pumping the UP to avoid complications of hot (i.e., highly excited) vibrational states when exciting the LP mode (*23*). The interpretation of these peaks was discussed in our previous work (*20*). Basically, the polariton transitions at ω_{MP} and ω_{LP} overlap with the 1→2 transition of the a_2'' and e′ modes, respectively [this assignment is further confirmed by spectral simulations (SM section 2.6) and input-output theory (SM section 4)]. Upon exciting the UP, when the waiting time was beyond the polariton lifetime, the [ω_{UP}, ω_{MP}] and [ω_{UP}, ω_{LP}] peaks corresponded, respectively, to the excited state population of the a_2'' and e′ modes. Therefore, the dynamics of these peaks reported the energy transfer between the a_2'' and e′ modes that was facilitated by pseudorotation and IVR. At a first glance, the dynamics of energy transfer under VSC (reported by [ω_{UP}, ω_{MP}] and [ω_{UP}, ω_{LP}] peak dynamics) were a bit faster than those of outside the cavity (reported by the [2022, 2010] diagonal and [2022, 1986] cross-peak dynamics) (Fig. 1D).

To quantify the energy-exchange dynamics upon pumping the UP, we used the kinetic model shown in Fig. 2A. First, the population of the UP relaxed to both the a_2'' and e′ dark modes within the polariton lifetime. Then, the a_2'' and e′ dark modes exchanged energy, through pseudorotation and IVR, and at the same time dissipated energy to their environment. The solution of this kinetic model provided a good fit to the [ω_{UP}, ω_{MP}] and [ω_{UP}, ω_{LP}] dynamics (Fig. 2B). Furthermore, the measured dynamics could be separated into three components: polariton relaxation to dark modes at short times (cavity leakage is implicitly accounted for; see SM section 2.3.2), energy exchange at intermediate times, and vibrational decay at long times. From the fitted results, k_{ex} under VSC was 0.113 ± 0.009 ps^{-1} at room temperature, 30% faster than that outside the cavity.

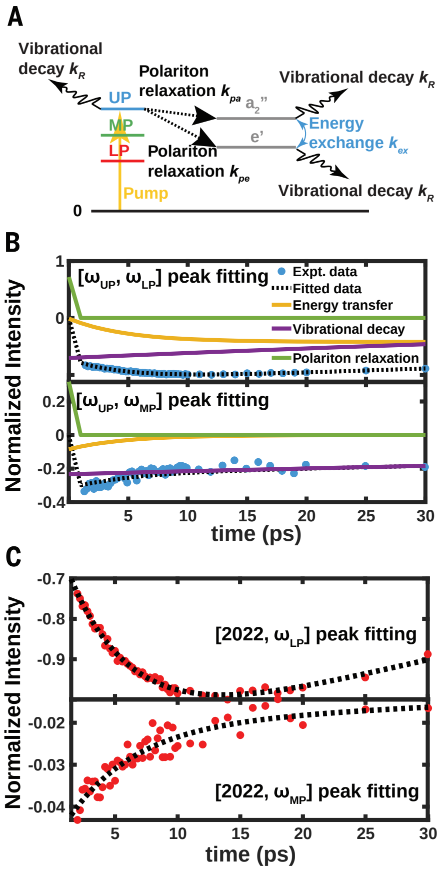


Fig. 2. Energy-exchange dynamics between a_2'' and e′ modes. (**A**) Schematic drawing of the kinetic model for $Fe(CO)_5$ under VSC. See SM section 2.3.2 for details of the kinetic model. k_{pe} (k_{pa}), the rate constant for the polariton relaxation from UP to the e′ (a_2'') modes; k_{ex}, the energy transfer rate constant between the e′ and a_2'' modes; k_R, the rate constant for the vibrational decay. (**B**) Experimental data (blue dots) and fits (black dotted lines) including each component for [ω_{UP}, ω_{LP}] (top) and [ω_{UP}, ω_{MP}] (bottom) peaks. (**C**) Experimental data (red dots) and fits (black dotted lines) for the [2022, ω_{LP}] (top) and [2022, ω_{MP}] (bottom) peaks when the a_2'' dark modes are pumped.

Using 2D IR and the same analysis, we found when exciting the a_2'' dark reservoir modes directly, the energy-exchange dynamics had similar trends as those outside the cavity (Fig. 2C) and k_{ex} was 0.090 ± 0.006 ps^{-1}. Clearly, VSC was only modifying the dynamics when the polaritons were pumped, whereas pumping the a_2'' dark reservoir modes caused the system to evolve similarly to the molecules outside the cavity, agreeing with the reservoir's purely molecular character. Similar findings have been predicted by a recent theoretical work (*24*). The contrast of dynamics between pumping the UP and dark a_2'' modes suggests that the energy-exchange rates depend on the initial populated states.

Although we have shown that VSC leads to faster energy exchange between the a_2'' and e′ modes, this acceleration could be due to an enhancement of either pseudorotation or IVR. To qualitatively distinguish between the two processes, we measured the vibrational anisotropy (*25*) associated with the cross-peaks. IVR involved energy transfer between e′ and a_2'' modes that were perpendicular to each other (Fig. 3A), and pseudorotation caused energy exchange between e′ and a_2'' modes that were parallel to each other (Fig. 3B). Thus, the anisotropy should start from −0.2 and 0.4 for the former and latter (*25*, *26*), respectively. In general, both processes occurred concurrently, and the initial anisotropy was between these values.

Outside the cavity, the initial value of anisotropy was ~0.06 for exciting the a_2'' modes (Fig. 3C), suggesting that pseudorotation dominated over IVR. However, under VSC, the opposite trend was observed: The anisotropy started at about −0.08 for exciting the UP (Fig. 3D). This contrast indicates that under VSC, IVR dominated over pseudorotation. Not surprisingly, when pumping the a_2'' dark modes under VSC, the anisotropy was ~0.06 (fig. S18), similar to the cavity-free case.

To determine the rate constants of pseudorotation and IVR, a more-detailed kinetic model was developed (see SM section 2.4) (*25*, *26*). The anisotropy can be calculated based on the energy-exchange dynamics simulated from the kinetic model. By fitting the measured anisotropy dynamics to the kinetic model, we determined the rate constants for pseudorotation (k_{ps}) and IVR (k_{IVR}) to be 0.035 ± 0.001 and 0.024 ± 0.001 ps^{-1}, respectively, outside the

cavity (Fig. 3C) and 0.022 ± 0.005 and 0.043 ± 0.002 ps^{-1} under VSC (Fig. 3D). The energy-exchange rate is $k_{ex} = k_{ps} + 2*k_{IVR}$ because the pseudorotation adiabatically morphs the a_2'' mode into one of the e′ modes but the IVR causes relaxation of a_2'' into both e″ modes (see SM, section 2.4.1e). The former results qualitatively agree with previous work showing that pseudorotation dominated the dynamics outside a cavity (*16*), except that now we quantified the relative contribution of IVR. The quantitative results agreed with the qualitative analysis above, indicating that VSC shifted the balance between pseudorotation and other energy-exchange channels: Outside the cavity, exciting the a_2'' mode yielded dynamics where pseudorotation dominated over IVR, and under VSC, exciting the UP promoted IVR and suppressed pseudorotation; yet, under VSC, exciting the a_2'' dark modes did not change the dynamics relative to molecules outside the cavity. We note that this effect is VSC-exclusive, because weak coupling to the cavity did not lead to the modification of the dynamics (see SM section 3.4); further, the acceleration of energy transfer and promotion of IVR through VSC was robust against different solvent environments (see results for 1-octanol, SM section 3.9).

The sharp contrast between the VSC dynamics starting in the UP and a_2'' dark reservoir modes is interesting because, even when the UP was excited, the population relaxed to the dark modes on a much shorter time scale than pseudorotation and IVR. The difference then lies in the relaxation processes available to the initial states. Several mechanisms could explain the faster IVR upon pumping the UP. For example, the decay from the UP to the a_2'' dark modes was accompanied by the excitation of low-frequency vibrations (i.e., phonons), and some of these phonons could be further excited during the energetically downhill IVR from a_2'' to e′ modes. It follows that IVR would be accelerated by the first scattering process, and this enhancement would not occur if the system were initialized in the a_2'' dark modes. A limitation of this hypothesis is that at room temperature, the phonons should have a high occupation number (≈10), which should not substantially change when the UP relax to dark modes (through one- or few-phonon excitation). Another possibility is that the VSC-induced speedup in the dynamics reflects a polariton-induced intermolecular vibrational energy transfer. In this case, though, the observed anisotropy dynamics should practically be zero, or have a fast decay (*20*), instead of starting from a negative value, owing to the lack of orientational correlation between donor and acceptor molecules. Conversely, the mild suppression of pseudorotation is surprising, because it is conventionally thought that these high-frequency vibrational modes do not drive the reaction. However, the coexistence of IVR enhancement and pseudorotation suppression suggests otherwise. By quickly going through IVR, molecules may lose their driving force for pseudorotation, leading to the pseudorotation slowdown. It is also possible that the pseudorotation motion is hindered by the phonons that are excited by the transition from the UP to the dark modes. The temperature-dependent measurements further showed that VSC shifts the thermodynamic parameters of activation in the same direction (see SM section 2.4), which has also been observed—and rather consistently—in reports of reaction kinetics altered by VSC (*27*). This correspondence supports that the insights obtained here should be relevant to understanding the previous experiments (*28*).

By using 2D IR to resolve ultrafast chemical dynamics with specific initial states, we quantified the energy-exchange dynamics in $Fe(CO)_5$ under VSC. We showed that when the UP was excited under VSC, IVR was promoted and pseudorotation was suppressed compared with the bare molecular system. However, pumping the dark reservoir modes under VSC led to little change in the dynamics compared with those outside the cavity. Because the population

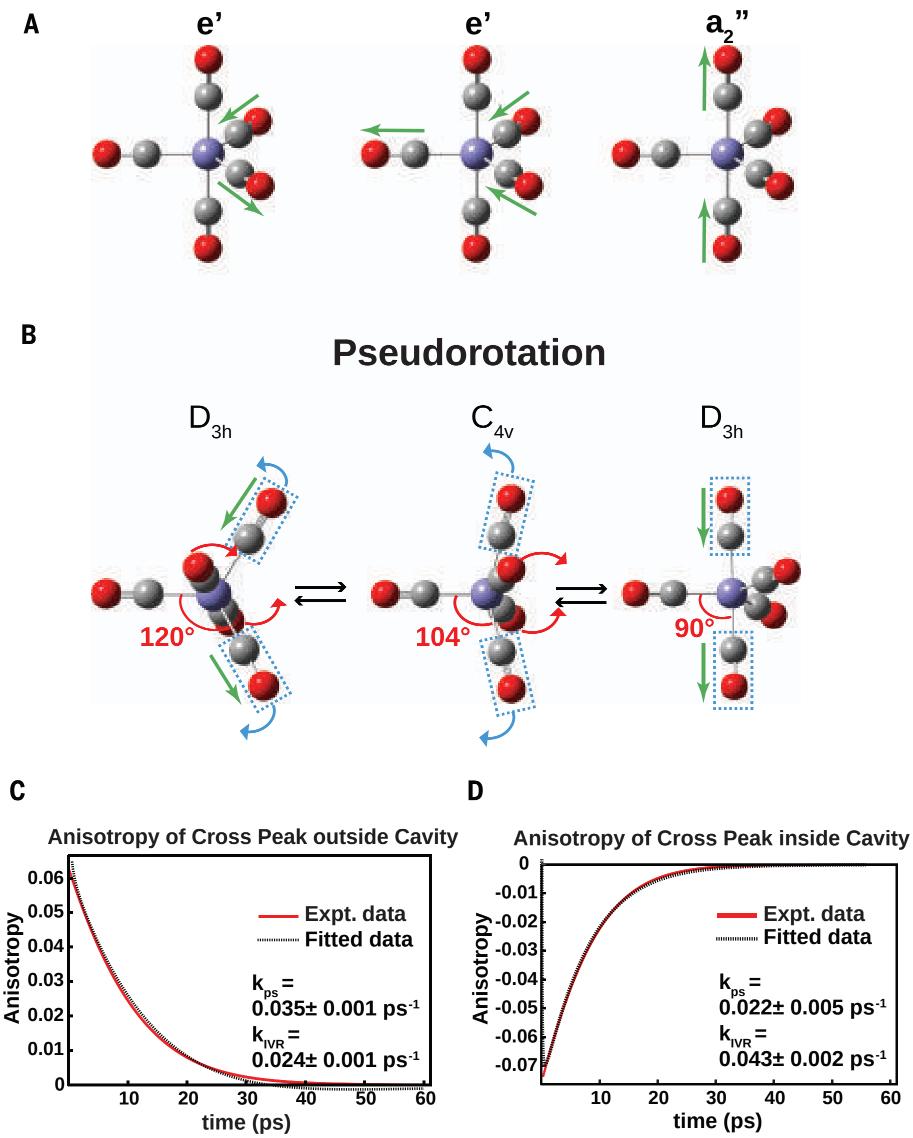


Fig. 3. Cross-peak anisotropy dynamics of IVR and pseudorotation. (**A**) Depiction of the eigenvectors for the a_2'' and doubly degenerate e′ vibrational modes of $Fe(CO)_5$. IVR leads to energy transfer between modes that are perpendicular to each other. (**B**) Pseudorotation leads to energy transfer between a_2'' and e′ modes that are parallel to each other. (**C** and **D**) Experimental anisotropy (red line) and corresponding fits (black dotted line) for the cross-peak of $Fe(CO)_5$ (C) outside the cavity ([2022, 1986]) and (D) under VSC ([ω_{UP}, ω_{MP}]). Listed are the rate constants, extracted from the fitting, of IVR (k_{IVR}) and pseudorotation (k_{ps}). The rate constants indicate that VSC accelerated IVR and suppressed pseudorotation.

at thermal equilibrium resides predominantly in the dark modes, the overall influence of VSC on $Fe(CO)_5$ should be negligible without external (e.g., laser) pumping. Yet, the present results show an important insight that unifies the works that report VSC-modified reactions and the ones that report or predict the opposite: Regardless of how reactions behave under thermally activated conditions, the basic concept of VSC-modified chemistry works—populated polaritons can influence chemical dynamics. These findings suggest that the future of VSC-modified thermal chemistry lies in controlling the dark modes, either by reducing the number of dark reservoir modes (*29*) or even going to the single-molecule regimes (*30*), for example, through cavity miniaturization or by making dark modes more delocalized through heterogeneity (*31*, *32*).

REFERENCES AND NOTES

1. J. P. Long, B. S. Simpkins, *ACS Photonics* **2**, 130–136 (2015).
2. A. Shalabney *et al*., *Nat. Commun.* **6**, 5981 (2015).
3. A. Thomas *et al*., *Angew. Chem. Int. Ed.* **55**, 11462–11466 (2016).
4. A. Thomas *et al*., *Science* **363**, 615–619 (2019).
5. F. J. Garcia-Vidal, C. Ciuti, T. W. Ebbesen, *Science* **373**, eabd0336 (2021).
6. J. Galego, C. Climent, F. J. Garcia-Vidal, J. Feist, *Phys. Rev. X* **9**, 021057 (2019).
7. T. E. Li, A. Nitzan, J. E. Subotnik, *J. Chem. Phys.* **152**, 234107 (2020).
8. I. Vurgaftman, B. S. Simpkins, A. D. Dunkelberger, J. C. Owrutsky, *J. Phys. Chem. Lett.* **11**, 3557–3562 (2020).
9. M. Du, J. A. Campos-Gonzalez-Angulo, J. Yuen-Zhou, *J. Chem. Phys.* **154**, 084108 (2021).
10. M. V. Imperatore, J. B. Asbury, N. C. Giebink, *J. Chem. Phys.* **154**, 191103 (2021).
11. X. Li, A. Mandal, P. Huo, *Nat. Commun.* **12**, 1315 (2021).
12. J. A. Campos-Gonzalez-Angulo, J. Yuen-Zhou, *J. Chem. Phys.* **156**, 194308 (2022).
13. J. A. Campos-Gonzalez-Angulo, J. Yuen-Zhou, *J. Chem. Phys.* **152**, 161101 (2020).
14. G. D. Wiesehan, W. Xiong, *J. Chem. Phys.* **155**, 241103 (2021).
15. R. S. Berry, *J. Chem. Phys.* **32**, 933–938 (1960).
16. J. F. Cahoon, K. R. Sawyer, J. P. Schlegel, C. B. Harris, *Science* **319**, 1820–1823 (2008).
17. F. H. Westheimer, *Acc. Chem. Res.* **1**, 70–78 (1968).
18. P. Törmä, W. L. Barnes, *Rep. Prog. Phys.* **78**, 013901 (2014).
19. B. Xiang *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4845–4850 (2018).
20. B. Xiang *et al*., *Science* **368**, 665–667 (2020).
21. B. Xiang, J. Wang, Z. Yang, W. Xiong, *Sci. Adv.* **7**, eabf6397 (2021).
22. B. Xiang, W. Xiong, *J. Chem. Phys.* **155**, 050901 (2021).
23. B. Xiang *et al*., *J. Phys. Chem. A* **123**, 5918–5927 (2019).
24. D. Wellnitz, G. Pupillo, J. Schachenmayer, *Commun. Phys.* **5**, 120 (2022).
25. R. M. Hochstrasser, *Chem. Phys.* **266**, 273–284 (2001).
26. A. Tokmakoff *et al*., *J. Chem. Phys.* **102**, 3919–3931 (1995).
27. K. Hirai, J. A. Hutchison, H. Uji-I, *ChemPlusChem* **85**, 1981–1988 (2020).
28. A. D. Dunkelberger, B. S. Simpkins, I. Vurgaftman, J. C. Owrutsky, *Annu. Rev. Phys. Chem.* **73**, 429–451 (2022).
29. T. E. Li, A. Nitzan, J. E. Subotnik, *J. Chem. Phys.* **154**, 094124 (2021).
30. C. C. Schäfer, J. Flick, E. Ronca, P. Narang, A. Rubio, arXiv:2104.12429 [quant-ph] (2021).
31. G. D. Scholes, *Proc. R. Soc. London Ser. A* **476**, 20200278 (2020).
32. M. Du, J. Yuen-Zhou, *Phys. Rev. Lett.* **128**, 096001 (2022).
33. T.-T. Chen, M. Du, Z. Yang, J. Yuen-Zhou, W. Xiong, Cavity-enabled enhancement of ultrafast intramolecular vibrational redistribution over pseudorotation, Zenodo (2022); https://doi.org/10.5281/zenodo.7110883.

ACKNOWLEDGMENTS

We thank H. H. Bhakta for assisting with the Mathematica code. **Funding:** T.-T.C. is supported by National Science Foundation (NSF) grant DMR-1848215. Z.Y. is supported by Air Force Office of Scientific Research grant FA9550-21-1-0369. W.X. acknowledges general support for summer salaries from Alfred P. Sloan Foundation grant FG-2020-12845 and NSF grant CHE-2101988. M.D. is supported through the American Chemical Society Petroleum Research Fund grant 60968-ND6. J.Y.-Z. was supported by the US Department of Energy, Office of Science, Basic Energy Sciences, Condensed Phase and Interfacial Molecular Science (CPIMS) program under Early Career Research Program Award DE-SC0019188. **Author contributions:** W.X. conceived the original idea and supervised the overall research. T.-T.C., Z.Y., and W.X. designed the experiments. T.-T.C. and Z.Y. conducted the experimental work. T.-T.C., M.D., J.Y.-Z., and W.X. analyzed data. M.D. and W.X. developed the anisotropy model. M.D. and J.Y.-Z. proposed mechanisms and developed the spectral theory in SM section 4. T.-T.C., M.D., J.Y.-Z., and W.X. interpreted the results and wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to support the conclusions of the main text and supplementary materials have been uploaded to Zenodo (*33*).

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add0276
Materials and Methods
Supplementary Text
Figs. S1 to S45
Tables S1 to S7
References (*34*–*42*)

Submitted 16 May 2022; resubmitted 14 August 2022
Accepted 26 September 2022
10.1126/science.add0276


OPEN ACCESS
Pushing the Boundaries of Knowledge
As AAAS's first multidisciplinary, open access journal, Science Advances publishes research that reflects the selectivity of high impact, innovative research you expect from the Science family of journals, published in an open access format to serve a vast and growing global audience. Check out the latest findings or learn how to submit your research: science.org/journal/sciadv
Science Advances
AAAS
GOLD OPEN ACCESS, DIGITAL, AND FREE TO ALL READERS

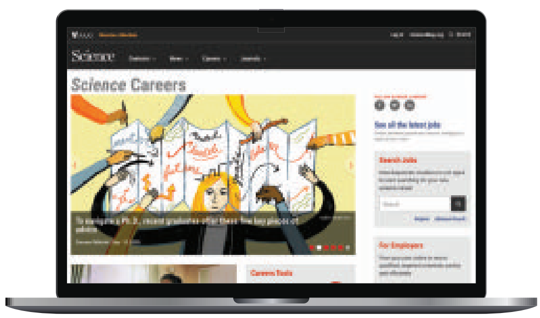
ScienceCareers online @sciencecareers.org



Science Careers helps you advance your career. Learn how !

- Register for a free online account on **ScienceCareers.org**.
- Search hundreds of job postings and find your perfect job.
- Sign up to receive e-mail alerts about job postings that match your criteria.
- Upload your resume into our database and connect with employers.
- Watch one of our many webinars on different career topics such as job searching, networking, and more.
- Download our career booklets, including Career Basics, Careers Beyond the Bench, and Developing Your Skills.
- Complete an interactive, personalized career plan at "my IDP."
- Visit our Employer Profiles to learn more about prospective employers.
- Read relevant career advice articles from our library of thousands.

Visit **ScienceCareers.org** today — all resources are free



ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

SCIENCECAREERS.ORG

NATIONAL ACADEMIES Sciences Engineering Medicine

POSTDOCTORAL RESEARCH ASSOCIATES
NRC RESEARCH ASSOCIATESHIP PROGRAMS

The National Academies of Sciences, Engineering, and Medicine administers postdoctoral and senior research awards at participating federal laboratories and affiliated institutions at locations throughout the U.S and abroad.

We are seeking highly qualified candidates who hold, or anticipate earning, a doctorate in a variety of fields of science or engineering. Degrees from foreign universities should be equivalent in training and research experience to a doctoral degree from a U.S. institution. Citizenship eligibility varies among the sponsoring laboratories.

Application deadline dates (four annual review cycles):
- February 1
- May 1
- August 1
- November 1

Awardees have the opportunity to:
- Conduct independent research in an area compatible with the interests of the sponsoring laboratory
- Devote full-time effort to research and publication
- Access the excellent and often unique facilities of the federal research enterprise
- Collaborate with leading scientists and engineers at the sponsoring laboratories

Awardee benefits include:
- Stipends ranging from $45,000 to $97,490; may be higher based on experience
- Health insurance (including dental and vision), relocation benefits, and a professional travel allowance

For detailed program information, to search Research Opportunities, and to contact prospective Research Adviser(s) visit **www.nas.edu/rap**.

NATIONAL ACADEMIES Sciences Engineering Medicine

POSTDOCTORAL RESEARCH FELLOWS
AIR FORCE SCIENCE & TECHNOLOGY FELLOWSHIP PROGRAM

The National Academies of Sciences, Engineering, and Medicine administers postdoctoral and senior fellowship awards at the U.S. Air Force Research Laboratory (AFRL), the U.S. Air Force Institute of Technology (AFIT), and the U.S. Air Force Academy (USAFA) under the Air Force Science & Technology Fellowship Program (AF STFP).

We are seeking highly qualified candidates who are U.S. citizens and hold, or anticipate earning, a doctorate in a variety of fields of science or engineering.

Application deadline dates (four annual review cycles):
- February 1
- May 1
- August 1
- November 1

Awardees have the opportunity to:
- Conduct independent research in an area compatible with the interests of the Air Force laboratories
- Devote full-time effort to research and publication
- Access the excellent and often unique Air Force research facilities
- Collaborate with leading scientists and engineers

Awardee benefits:
- Base stipend starting at $76,542; may be higher based on experience
- Health insurance (including dental and vision), relocation benefits, and professional travel allowance

For detailed program information, to search for AFRL, AFIT, and USAFA Research Opportunities, and to contact prospective Research Adviser(s), visit **www.nas.edu/afstfp**.



University of Missouri

The University of Missouri (MU) Department of Biochemistry invites applications for a **tenure-track faculty position in plant radiochemistry**. We seek candidates working at the interface of plant metabolism and radiochemistry that leverage the generation of radio-isotopes at the University of Missouri Research Reactor (MURR). The faculty member is expected to develop a robust and externally funded research program exploiting the resources and-, capabilities; of MURR to address fundamental questions in plant biology. Examples include, studies of plant carbon or nitrogen dynamics utilizing radioactive 11C or 13N, micronutrient assimilation and/or exchange between plants, soil, and microorganisms, as well as metabolic ¯ux and regulation using isotope ratio analyses. Applications at the Assistant Professor level are encouraged. Candidates at a higher level will be considered if there is a record of extramural research funding, publications, teaching and service expected of tenured faculty at a research-intensive university.

The department has an undergraduate and graduate teaching mission in the College of Agriculture and Natural Resources (CAFNR). The successful candidate will be a member of the Interdisciplinary Plant Group (IPG), a world-renowned group of researchers from seven departments/units with interests in plant biology. He/she will also be a core faculty member of MURR, which operates the highest power (10 MW) university research reactor in the US and is planning a further expansion. For additional information, visit **https://biochem.missouri.edu** and **https://ipg.missouri.edu**.

APPLICATION: To apply, visit **https://biochem.missouri.edu/open-positions/**, click on **Open Positions** and submit: **(1)** a cover letter; **(2)** curriculum vitae; **(3)** a one-page summary of research accomplishments with two pages of future plans; **(4)** a one-page narrative of teaching philosophy; **(5)** a description of current and future efforts to expand diversity & inclusion in science; and **(6)** names and contact information of four references. Review of applications will begin on **December 15, 2022** and will continue until the position is filled. For information, contact **biochemchair@missouri.edu**.

The University of Missouri is an Equal Access, Equal Opportunity, and Affirmative Action Employer. We are fully committed to achieving the goal of a diverse and inclusive academic community of faculty, staff and students. We seek individuals who are committed to this goal and our core campus values of respect, responsibility, discovery and excellence.

WORKING LIFE

By **Walter P. Suza**

Silent no longer

Fear kept me silent. When I was a child in Tanzania, kids from a different tribe would wait for me at the end of the school day to beat me up. The adrenaline rush of fear may have sometimes helped me outrun my nemeses. Yet it also kept me silent. I feared that if I told anyone at home, my older siblings would try to intervene, which would only make future beatings worse. That was many years ago. The United States is my country now. But I am still afraid. To America, I am not just a man; I am a Black man. And for years, the baggage of that label kept me afraid and silent.

Since coming to the United States as a graduate student, I have been subject to microaggressions and worse. On the first day of my biochemistry class, a white professor asked, "Are you sure you are in the right class?" as I was walking in the door. Outside a restaurant and sports bar off-campus, a white man called me the N-word. During my postdoctoral training, a white colleague giving a lab tour to a white family of a prospective student asked them, "Have you seen the Black guy?" I assume he did not see me sitting in the corner.



Despite the grave reality that such dehumanizing language can pave the way for death at the hands of the police and vigilantes, I was afraid to speak about it in public. Professionally, I was afraid if I spoke, potential employers might view me as a troublemaker and I would face greater barriers to pursuing my goal of becoming a tenured professor. Personally, I feared it might jeopardize my chances of being awarded permanent residency and eventually becoming a U.S. citizen. So, I kept my head down and focused on my work and my family.

"Despite the threats, I continue to write."

The stress from not speaking out felt constant and inescapable. I felt trapped inside my own skin.

I was afraid to speak about being followed in grocery stores. I was afraid to speak about a white woman at the park ordering her child not to play with my daughter. I was afraid to speak about being seated in the back of the restaurant when we were the only family there.

I remained silent after a white vigilante shot and killed Trayvon Martin. I remained silent after a white police officer shot and killed Michael Brown. I remained silent after another white police officer shot and killed Tamir Rice.

I moved on to a non–tenure-track faculty position, pleased to be progressing in my career. But I could not outrun racism. On campus, incidents of white supremacist symbols and racist graffiti were frequently reported. Off campus, I saw the Confederate flag being flown.

Still, I remained silent, unable to overcome my fear. Instead of speaking out, I dove into diversity and inclusion work. I sought professional development to learn how to make my classroom more inclusive, hosted a student from an internship program intended in part to increase diversity, and served on department committees to strengthen diversity, equity, inclusion, and belonging. I also read about the history of the U.S. civil rights struggle.

Over time, these efforts helped me gain the courage to write about racial and social justice for the student newspaper. Several students and colleagues offered encouraging, supportive responses, which helped reduce my fear of career repercussions. But I was still afraid to speak out to the public at large.

The tipping point came when Ahmaud Arbery was shot dead by white men and George Floyd was choked to death by a white police officer. My rage and despair were nearly overwhelming. I was still afraid. But by that time I had become a U.S. citizen, and I felt I could no longer stay silent. I owed it to myself, and to those who were more vulnerable, to speak out.

So, I started to write opinion columns for local newspapers. Some readers encourage me to keep writing. Some tell me to "get out of my state." Some make phone calls to my work, calling me a racist. Some send disturbing threats to my home and work. Despite the threats, I continue to write.

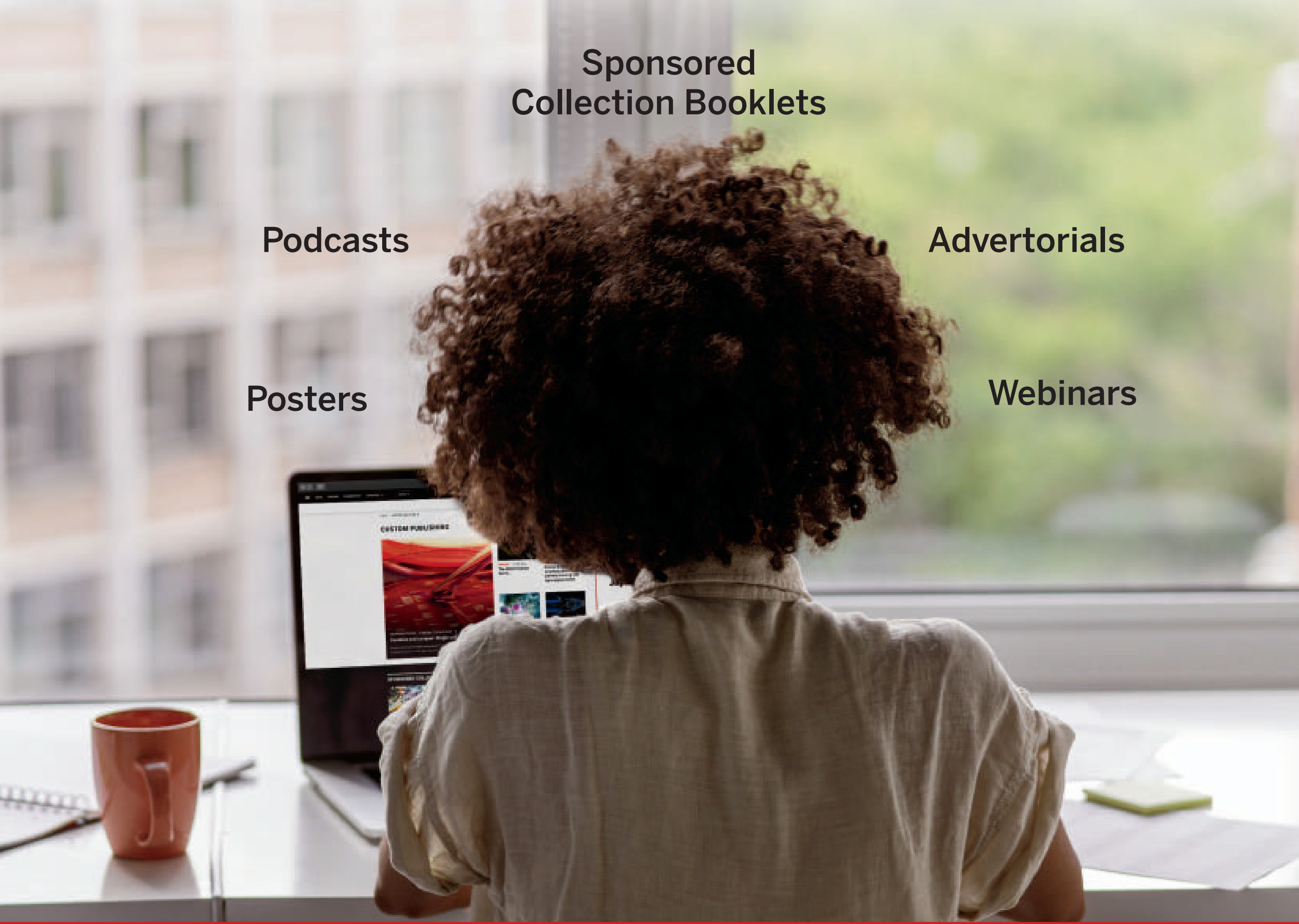
I am still a Black man in America, with all the resulting challenges and fears. I am still afraid. But I have learned that fear need not mean silence. Instead, I can channel the adrenaline—not to outrun and escape my fear, but to face it head on, in the hopes that someday I and others like me will no longer have cause to feel afraid. ■

Walter P. Suza is an adjunct associate professor at Iowa State University.
Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

QUALITY CONTENT FOR THE GLOBAL SCIENTIFIC COMMUNITY

Multiple ways to stay informed on issues related to your research



Science
AAAS

Scan the code and start exploring the latest advances in science and technology innovation!

Science.org/custom-publishing

Brought to you by the *Science*/AAAS Custom Publishing Office.

Posters **Podcasts** **Sponsored Collection Booklets** **Advertorials** **Webinars**

TRILLIONS OF MICROBES

ONE ESSAY

The NOSTER *Science* Microbiome Prize is an international prize that rewards innovative research by investigators who have completed their terminal degree in the last 10 years and are working on the functional attributes of microbiota. The research can include any organism with the potential to contribute to our understanding of human or veterinary health and disease, or to guide therapeutic interventions. The winner and finalists will be chosen by a committee of independent "scientists" chaired by a senior editor of *Science*. The top prize includes a complimentary membership to AAAS, an online subscription to *Science*, and USD 25,000. Submit your research essay today.



Jennifer Hill, Ph.D.
2022 Winner

NOSTER | Science
MICROBIOME
PRIZE

Apply by 24 January 2023 at www.science.org/noster

Sponsored by Noster Inc